

# THE ANNALS of MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. G. CARVER, 1900-1933

EDITED BY S. S. WILKS, 1933-1949

THE OFFICIAL JOURNAL OF THE INSTITUTE  
OF MATHEMATICAL STATISTICS

## Contents

	PAGE
An Approach to Time Series Analysis.....	EMANUEL PARZEN 951
Some Model I Problems of Selection.....	E. L. LEHMANN 990
Bayes Rules for a Common Multiple Comparisons Problem and Related Student- <i>t</i> Problems.....	DAVID B. DUNCAN 1013
The Use of Least Favorable Distributions in Testing Composite Hypotheses.....	H. E. REINHARDT 1034
Asymptotic Efficiency in Polynomial Estimation.....	PAUL G. HOEL 1042
Maximum Likelihood Estimation of a Linear Functional Relationship	C. VILLEGAS 1048
Sequential $\chi^2$ - and $T^2$ -Tests...J. EDWARD JACKSON AND RALPH A. BRADLEY	1063
Estimating the Parameters of Negative Exponential Populations from One or Two Order Statistics.....	H. LEON HARTER 1078
On the Two Sample Problem: A Heuristic Method for Constructing Tests	V. P. GODAMBE 1091
A Nonparametric Test for the Problem of Several Samples.	V. P. BHAPKAR 1108
Distribution of the Anderson-Darling Statistic.....	PETER A. W. LEWIS 1118
Errors in Discrimination.....	S. JOHN 1125
On the Monotonic Character of the Power Functions of Two Multivariate Tests.....	S. N. ROY AND W. F. MIKHAIL 1145
The Moments of Elementary Symmetric Functions of the Roots of a Matrix in Multivariate Analysis.....	TITO A. MIJARES 1152
Variance Components in the Unbalanced 2-Way Nested Classification	S. R. SEARLE 1161
Some Main-Effect Plans and Orthogonal Arrays of Strength Two	SIDNEY ADDELMAN AND OSCAR KEMPTHORNE 1167
On A Geometrical Method of Construction of Partially Balanced Designs with Two Associate Classes.....	ESTHER SEIDEN 1177
On Some Methods of Construction of Partially Balanced Arrays	I. M. CHAKRAVARTI 1181
Some Further Designs of Type O:PP.....	G. H. FREEMAN 1180
Sufficiency in the Undominated Case.....	D. L. BURKHOLDER 1191

(Continued on back cover)

# THE ANNALS OF MATHEMATICAL STATISTICS

**Subscription Rates.** Current issues are \$15 per volume (four issues of one calendar year) in the U. S. and Canada, \$12 per volume elsewhere. Single issues are \$4. Back numbers for all issues up to and including 1956 (Vol. XXVII) are \$12 per volume, \$3.50 per issue; issues from 1957 on are \$15 per volume, \$4 per issue; for complete sets, \$200 for the first 25 volumes, \$10 per additional volume purchased at the same time as Volumes I through XXV.

Rates to members of the Institute of Mathematical Statistics are lower (see inside back cover).

Communications concerning subscriptions, back numbers, payment of dues, membership, changes of address, etc., should be addressed to Gerald Lieberman, Treasurer, Institute of Mathematical Statistics, Department of Statistics, Stanford University, Stanford, Calif.

**Editorial Office,** Department of Statistics, University of California, Berkeley 4, California. J. L. Hodges, Jr. Editor. (See note on next page about change of Editor.)

**Preparation of manuscripts.** Manuscripts should be submitted to the editorial office. Each manuscript should be typewritten, double spaced, with wide margins at sides, top, and bottom, and the original should be submitted with one additional copy, on paper that will take corrections. Dittoed or mimeographed papers are acceptable only if completely legible. Footnotes should be reduced to a minimum, and where possible replaced by remarks in the text, or a bibliography at the end of the paper; formulae in footnotes should be avoided. References should follow current *Annals* style, and should be numbered alphabetically according to authors' names.

Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed.

Authors are asked to keep in mind the typographical difficulties of complicated mathematical formulae. The difference between capital and lower-case letters should be clearly shown; care should be taken to avoid confusion between such pairs as zero and the letter O, the numeral 1 and the letter l, numeral 1 used as superscript and prime ('), alpha and a, kappa and k, mu and u, nu and v, eta and n, etc. Subscripts or superscripts should be clearly below or above the line. Bars above groups of letters (e.g.,  $\log \bar{x}$ ) and underlined letters (e.g.,  $\bar{x}$ ) are difficult to print and should be avoided. Symbols are automatically italicized by the printer and should not be underlined on manuscripts. Boldface letters may be indicated by underlining with a wavy line on the manuscript; boldface subscripts and superscripts are not available. Complicated exponentials should be represented with the symbol exp. In writing square roots the fractional exponent is preferable to the radical sign. Fractions in the body of the text (as opposed to displayed expressions) and fractions occurring in the numerators or denominators of fractions are preferably written with the solidus; thus  $(a + b)/(c + d)$  rather than  $\frac{a + b}{c + d}$ .

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers will be furnished at cost.

---

Mail to the *Annals of Mathematical Statistics* should be addressed to either the Editor or the Treasurer, as described above. It should not be addressed to Waverly Press.

---

COMPOSED AND PRINTED AT THE  
WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Second-class postage paid at Baltimore, Maryland



## EDITORIAL STAFF

### EDITOR

J. L. HODGES, JR.

### ASSOCIATE EDITORS

DOUGLAS G. CHAPMAN  
A. P. DEMPSTER  
ULF GRENANDER

SAMUEL KARLIN  
OSCAR KEMPTHORNE  
J. KIEFER

E. L. LEHMANN  
LEONARD J. SAVAGE  
DAVID L. WALLACE

### WITH THE COOPERATION OF

PATRICK BILLINGSLEY  
J. R. BLUM  
R. C. BOSE  
ROBERT J. BUEHLER  
D. L. BURKHOLDER

CYRUS DERMAN  
J. L. DOOB  
MEYER DWASS  
D. A. S. FRASER

EUGENE LUKACS  
INGRAM OLKIN  
J. W. PRATT  
RONALD PYKE

HOWARD RAIFFA  
WALTER L. SMITH  
LIONEL WEISS  
R. A. WIJSMAN

### PAST EDITORS OF THE ANNALS

H. C. CARVER, 1930-1938

E. L. LEHMANN, 1953-1955

S. S. WILKS, 1938-1949

T. E. HARRIS, 1955-1958

T. W. ANDERSON, 1950-1952

WILLIAM KRUSKAL, 1958-1961

Published quarterly by the Institute of Mathematical Statistics in March,  
June, September and December.

**CHANGE OF EDITOR.** Manuscripts should be submitted to the present Editor, J. L. Hodges Jr., Department of Statistics, University of California, Berkeley 4, California. Correspondence about manuscripts submitted before July 1, 1961 should be directed to the Past Editor, W. H. Kruskal.

## IMS INSTITUTIONAL MEMBERS

ABERDEEN PROVING GROUNDS, BALLISTIC RESEARCH LABORATORIES, Aberdeen, Maryland  
AEROJET-GENERAL CORPORATION, P. O. Box 296, Azusa, California  
AMERICAN VISCOSSE CORPORATION, MARCUS HOOK, Pennsylvania  
ATLANTIC REFINING COMPANY, 2700 Passyunk Avenue, Philadelphia, Pa.  
BELL TELEPHONE LABORATORIES, INC., TECHNICAL LIBRARY, 463 West Street, New York 14,  
New York  
BENDIX AVIATION CORPORATION, 1200 Fisher Bldg., Detroit, Michigan  
BOEING AIRPLANE COMPANY, Box 3707, Seattle, Washington  
CALIFORNIA RESEARCH CORPORATION, P. O. Box 1627, Richmond, California  
CASE INSTITUTE OF TECHNOLOGY, STATISTICAL LABORATORY, Cleveland 6, Ohio  
CATHOLIC UNIVERSITY OF AMERICA, STATISTICAL LABORATORY, MATHEMATICS DEPARTMENT,  
Washington, D. C.  
COLUMBIA UNIVERSITY, DEPARTMENT OF MATHEMATICAL STATISTICS, New York 27, N. Y.  
CORNELL UNIVERSITY, MATHEMATICS DEPARTMENT, Ithaca, New York  
FORD MOTOR COMPANY, P. O. Box 2053, Dearborn, Michigan  
GENERAL ELECTRIC COMPANY, Building C37, Room 248, Schenectady, New York  
GENERAL MOTORS CORPORATION, RESEARCH LABORATORIES, Twelve Mile and Mound  
Roads, Warren, Michigan  
INDIANA UNIVERSITY, THE LIBRARY, Bloomington, Indiana  
INTERNATIONAL BUSINESS MACHINES CORPORATION, MATHEMATICS AND APPLIED SCIENCE  
LIBRARY, 1271 Avenue of the Americas, New York 20, N. Y.  
IOWA STATE UNIVERSITY, STATISTICAL LABORATORY, Ames, Iowa  
LOCKHEED AIRCRAFT CORPORATION, ENGINEERING LIBRARY, Burbank, California  
MICHIGAN STATE UNIVERSITY, DEPARTMENT OF STATISTICS, East Lansing, Michigan  
MINNESOTA MINING AND MANUFACTURING COMPANY, APPLIED MATHEMATICS AND STA-  
TISTICS, St. Paul, Minnesota  
MONSANTO CHEMICAL COMPANY, 800 North Lindbergh Blvd., St. Louis 66, Missouri

(Continued on next page)

NATIONAL CASH REGISTER COMPANY, RESEARCH DEPARTMENT, Main and K Streets, Dayton 9, Ohio

NATIONAL SECURITY AGENCY, Fort George G. Meade, Maryland

NORTHWESTERN UNIVERSITY, DEPARTMENT OF MATHEMATICS, Evanston, Illinois

PRINCETON UNIVERSITY, DEPARTMENT OF MATHEMATICS, SECTION OF MATHEMATICAL STATISTICS, Princeton, New Jersey

PURDUE UNIVERSITY LIBRARIES, Lafayette, Indiana

OKLAHOMA STATE UNIVERSITY, DEPARTMENT OF MATHEMATICS, Stillwater, Oklahoma

RADIO CORPORATION OF AMERICA, R.C.A. LABORATORIES LIBRARY, Princeton, New Jersey

RAMO-WOOLRIDGE CORPORATION, Los Angeles, California

REMINGTON RAND—UNIVAC DIVISION, 315 Park Avenue South, New York 10, N. Y.

SANDIA CORPORATION, Sandia Base, Albuquerque, New Mexico

SOUTHERN METHODIST UNIVERSITY, MATHEMATICS DEPARTMENT, Dallas 5, Texas

SPACE TECHNOLOGY LABORATORIES, P. O. Box 95001, Los Angeles 45, California

STANFORD UNIVERSITY, GIRSHICK MEMORIAL LIBRARY, Stanford, California

STATE UNIVERSITY OF IOWA, Iowa City, Iowa

UNION OIL COMPANY OF CALIFORNIA, UNION RESEARCH CENTER, Box 76, Brea, California

UNITED STATES STEEL CORPORATION LIBRARY, Monroeville, Penna.

UNIVERSITY OF CALIFORNIA, STATISTICAL LABORATORY, Berkeley, California

UNIVERSITY OF ILLINOIS, SERIALS DEPARTMENT, Urbana, Illinois

UNIVERSITY OF MICHIGAN, DEPARTMENT OF MATHEMATICS, Ann Arbor, Michigan

UNIVERSITY OF NORTH CAROLINA, DEPARTMENT OF STATISTICS, Chapel Hill, North Carolina

UNIVERSITY OF PUERTO RICO, SCHOOL OF TROPICAL MEDICINE, San Juan, Puerto Rico

UNIVERSITY OF WASHINGTON, LABORATORY OF STATISTICAL RESEARCH, Seattle, Washington

WESTINGHOUSE ELECTRIC CORPORATION, RESEARCH LABORATORIES, Pittsburgh 35, Pennsylvania

W. R. GRACE AND COMPANY, RESEARCH DIVISION, Washington Research Center, Clarks-ville, Maryland

W. R. GRACE AND COMPANY, DEWEY AND ALMY CHEMICAL DIVISION, 62 Whittemore Avenue, Cambridge 40, Massachusetts

## AN APPROACH TO TIME SERIES ANALYSIS<sup>1</sup>

BY EMANUEL PARZEN

*Stanford University*

**Summary.** It may fairly be said that modern time series analysis is a subject which embraces three fields which while closely related have tended to develop somewhat independently. These fields are (i) statistical communication and control theory, (ii) the probabilistic (and Hilbert space) theory of stochastic processes possessing finite second moments, and (iii) the statistical theory of regression analysis, correlation analysis, and spectral (or harmonic) analysis of time series. In this paper it is my aim to show the close relation between these fields and to summarize some recent developments.

The topics discussed are (i) stationary time series and their statistical analysis, (ii) prediction theory and the Hilbert space spanned by a time series, and (iii) regression analysis of time series with known covariance function. In particular, I describe a new approach to prediction and regression problems using reproducing kernel Hilbert spaces.

**1. Introduction.** A set of observations arranged chronologically is called a time series. Time series are observed in connection with quite diverse phenomena, and by a wide variety of researchers, such as (1) the economist observing yearly wheat prices, (2) the geneticist observing daily egg production of a certain breed of hen, (3) the meteorologist studying daily rainfall in a given city, (4) the physicist studying the ambient noise level at a given point in the ocean, (5) the aerodynamicist studying atmospheric turbulence gust velocities, (6) the electronic engineer studying the internal noise of a radio receiver, and so on.

Time series analysis constitutes one of the most important tools of the economist. Consider the prices or quantities of commodities traded on an exchange. The record of prices or quantities over time may be represented as a fluctuating function (or wiggly record). The analysis of such economic time series is a problem of great interest to economists desiring to explain the dynamics of economic systems and to speculators desiring to forecast prices.

Techniques of time series analysis have long been used in science and engineering (for example, to smooth data and to search for "periodicities" [6]). The theory and practice of time series analysis is assuming new importance in the space age since a wide variety of problems involving communication and/or control (involving such diverse problems as the automatic tracking of moving

---

Received September 27, 1960.

<sup>1</sup> An address presented on August 25, 1960 at the Stanford meetings of the Institute of Mathematical Statistics by invitation of the IMS Committee on Special Invited Papers. This paper was prepared with the partial support of the Office of Naval Research (Nonr-225-21). Reproduction in whole or part is permitted for any purpose of the United States Government.

objects, the reception of radio signals in the presence of natural and artificial disturbances, the reproduction of sound and images, the design of guidance systems, the design of control systems for industrial processes, and the analysis of any kind of record representing observation over time) may be regarded as problems in time series analysis.

To represent a time series, one proceeds as follows. The set of time points at which measurements are made is called  $T$ . The observation made at time  $t$  is denoted by  $X(t)$ . The set of observations  $\{X(t), t \in T\}$  is called a time series.

In regard to the index set  $T$ , there are cases of particular importance. One may be observing (i) a discrete parameter time series  $X(t)$ , in which case one assumes  $T$  is a finite set of points written  $T = \{1, 2, \dots, N\}$ , (ii) a continuous parameter time series, in which case  $T$  is a finite interval written  $T = \{t: 0 \leq t \leq L\}$ , (iii) a multiple (discrete or continuous parameter) time series  $\{X_1(t), \dots, X_k(t), t \in T'\}$  which may be written as a time series  $\{X(t), t \in T\}$  with index set  $T = \{(j, t): j = 1, \dots, k \text{ and } t \in T'\}$ , or (iv) a space field  $X(x, y, z, t)$  defined on space-time which is a function of three coordinates of position and one coordinate of time.

The basic idea of the statistical theory of analysis of a time series  $\{X(t), t \in T\}$  is to regard the time series as being an observation made on a family of random variables  $\{X(t), t \in T\}$ ; that is, for each  $t$  in  $T$ ,  $X(t)$  is an observed value of a random variable. A family of random variables  $\{X(t), t \in T\}$  is called a *stochastic process*. An observed time series  $\{X(t), t \in T\}$  is thus regarded as an observation (or, in a different terminology, a realization) of a stochastic process  $\{X(t), t \in T\}$ .

It has been pointed out by various writers (see, for example, Neyman [34]) that there are two broad categories of statistical problems: problems of stochastic model building for natural phenomena and problems of statistical decision making. These two categories of problems are well illustrated in the analysis of economic time series; some study time series in order to understand the mechanism of the economic system while others study time series with the simple aim of being able to forecast, for example, stock market prices. In general, it may be said that the aims of time series analysis are

- (1) to understand the mechanism generating the time series,
- (2) to predict the behavior of the time series in the future. To attack either of these problems, one adopts a model for the time series.

A model often adopted for the analysis of an observed time series  $\{X(t), t \in T\}$  is to regard  $X(t)$  as the sum of two functions:

$$(1.1) \quad X(t) = m(t) + Y(t), \quad t \in T.$$

We call  $m(t)$  the mean value function and  $Y(t)$  the fluctuation function.

The stochastic process  $Y(t)$  is assumed to possess finite second moments, and to have zero means and *covariance kernel*

$$(1.2) \quad K(s, t) = E[Y(s)Y(t)].$$

In addition it is often assumed that  $Y(t)$  is a *normal process* in the sense that



for every finite subset  $\{t_1, \dots, t_n\}$  of  $T$ , the random variables  $Y(t_1), \dots, Y(t_n)$  are jointly normally distributed.

The mean value function

$$(1.3) \quad m(t) = E[X(t)]$$

is assumed to belong to a known class  $M$  of functions. For example,  $M$  may be the set of all linear combinations of  $q$  known functions  $w_1(t), \dots, w_q(t)$ ; then, for  $t$  in  $T$ ,

$$(1.4) \quad m(t) = \beta_1 w_1(t) + \dots + \beta_q w_q(t),$$

for some coefficients  $\beta_1, \dots, \beta_q$  to be estimated. Other possible assumptions often made concerning the mean value function  $m(t)$  are as follows: (i)  $m(t)$  represents a *systematic oscillation*,

$$(1.5) \quad m(t) = \sum_{j=1}^q A_j \cos(\omega_j t + \varphi_j)$$

in which the amplitudes  $A_j$ , the angular frequencies  $\omega_j$ , and the phases  $\varphi_j$  are constants, some of which are given and the rest are unknown and are to be estimated; (ii)  $m(t)$  represents a *polynomial trend*,

$$(1.6) \quad m(t) = \sum_{j=0}^{q-1} \beta_j t^j,$$

an assumption often adopted if  $m(t)$  represents the trajectory [given by  $m(t) = x_0 + vt + \frac{1}{2}at^2$ , say] of a moving object, or (iii)  $m(t)$  is the sum of a systematic oscillation and a polynomial trend, an assumption traditionally adopted in treating economic time series.

Early workers in time series analysis sought to explain the dependence between successive observations of a time series  $X(t)$  by assuming that  $X(t)$  [sometimes written  $X_t$ ] was generated by a scheme of the following kind:

$$(1.7) \quad X_t = m(t) + Y_t$$

where  $m(t)$  represents a systematic oscillation of the form of (1.5) and the fluctuations  $Y_1, \dots, Y_n$  are assumed to be independent, normal random variables with mean 0 and common unknown variance  $\sigma^2$ .

The model given by (1.7) is called the scheme of *hidden periodicities* and was first introduced by Schuster ([47], [48]). The method used to estimate the frequencies  $\omega_j$  (or, equivalently, the periods  $2\pi/\omega_j$ ) is called *periodogram analysis*. The problem of tests of significance in periodogram analysis ([11], [18]) played an important role in the early history of time series analysis.

Approaches to time series analysis which seem to be more fruitful than periodogram analysis (see Kendall [22]) were pioneered by Yule and Slutsky in the 1920's.

Yule's researches [60] led to the notion of the autoregressive scheme, in which a time series  $X_t$  is assumed to be generated as a linear function of its past values,

plus a random shock; in symbols, for some integer  $m$  (called the order of the autoregressive scheme) and constants  $a_1, \dots, a_m$

$$X_t = a_1 X_{t-1} + \dots + a_m X_{t-m} + \eta_t$$

in which the sequence  $\{\eta_t\}$  consists of independent identically distributed random variables. In particular, Yule showed that an autoregressive scheme of order 2 provided a better model for sunspots than did the scheme of hidden periodicities.

Slutsky's researches [51] led to the notion of a *moving average scheme*, in which a time series  $X_t$  is assumed to be generated as a finite moving average of a sequence of independent and identically distributed random variables  $\{\eta_t\}$ ; in symbols, for some integer  $m$  and constants  $a_0, \dots, a_m$

$$X_t = a_0 \eta_t + a_1 \eta_{t-1} + \dots + a_m \eta_{t-m}.$$

Slutsky showed that moving averages exhibit properties of disturbed periodicity and consequently can be used as a model for oscillatory time series. In particular, Slutsky proved the Sinusoidal Limit Theorem which showed that a sine wave could be approximated by a moving average scheme.

In the 1930's and 1940's, the probabilistic theory of stationary time series was developed, first as a result of the development of ergodic theory and then as a result of prediction theory. That the autoregressive and moving average schemes may be interpreted as special cases of the theory of stationary processes was pointed out by Wold [58] in 1938 (see [57], p. 169). Thus the link was established between the statistical theory of analysis of time series and the probabilistic theory of the structure of time series. In the last twenty years, an extensive literature has developed exploring this link.

It may fairly be said that modern time series analysis is a subject which embraces three fields which while closely related have tended to develop somewhat independently. These fields are (i) statistical communication and control theory ([26], [32]) (ii) the probabilistic (and Hilbert space) theory of stochastic processes possessing finite second moments ([9], Chaps. 9–12; [28], Chap. 10), and (iii) the statistical theory of regression analysis, correlation analysis, and spectral (or harmonic) analysis of time series ([13], [17], [23], [58]). In this paper it is my aim to show the close relation between these fields and to summarize some recent developments with which I have been closely associated. The contents of the paper are as follows.

(I) *Stationary time series and their statistical analysis.* While it is a fiction to regard an observed time series as having zero means, it is mathematically convenient to consider the analysis of time series under this assumption. Consequently, one may consider the analysis of an observed time series  $\{X(t), t \in T\}$  with vanishing mean value function and *unknown* covariance function.

It has long been traditional among physical scientists to regard time series as arising from a superposition of sinusoidal waves of various amplitudes, frequencies, and phases. In the theory of time series analysis and statistical communications theory, a central role is played by the notion of the spectrum of a

time series. For the time series analyst, the spectrum represents a basic tool for determining the mechanism generating an observed time series. For the communication theorist, the spectrum provides the major concept in terms of which to analyze the effect of passing stochastic processes (representing either signal or noise) through linear (and, to some extent, non-linear) devices. *Spectral* (or *harmonic*) *analysis* is concerned with the theory of the decomposition of a time series into sinusoidal components. For many time functions  $\{X(t), -\infty < t < \infty\}$ , such a decomposition is provided by the Fourier transform

$$S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} X(t) dt.$$

Unfortunately, no meaning can be attached to this integral for many stochastic processes  $\{X(t), -\infty < t < \infty\}$  since their *sample functions are nonperiodic undamped functions* and therefore do not belong to the class of functions dealt within the usual theories of Fourier series and Fourier integrals. Nevertheless, it is possible to define a notion of *harmonic analysis of stochastic processes* (that is, a method of assigning to each frequency  $\omega$  a measure of its contribution to the "content" of the process) as was first shown by N. Wiener [56] and A. Khintchine [24]. Among the stochastic processes which possess a harmonic analysis, *stationary* processes are most important since a time series may be represented as a superposition of sinusoidal waveforms with "independent amplitudes" if and only if it is stationary (see Section 4 for a more precise form of this assertion). In Section 2, some basic results concerning stationary processes are summarized.

Much of the recent statistical literature on time series analysis has been concerned with questions of statistical inference on stationary time series and especially with

(i) deriving the exact and asymptotic distributions of various estimates of the covariance functions  $R(v)$  and the normalized covariance (or correlation) function  $\rho(v) = R(v)/R(0)$  of a stationary time series,

(ii) fitting stationary time series by mechanisms (such as autoregressive schemes or moving average schemes) which are completely specified except for a finite number of parameters, and with estimating the parameters of such schemes,

(iii) estimating (and forming confidence sets) for the spectral density function and spectral distribution function of a stationary time series.

For many purposes, it is preferable to estimate the spectrum of a stationary time series rather than its correlation function, since many aspects of a stationary time series are best understood in terms of its spectrum. The spectrum enables one to (i) investigate the physical mechanism generating a time series, (ii) determine the behavior of a dynamic linear system in response to random excitations, and (iii) possibly simulate a time series. Other uses of the spectrum are as operational means (i) of transmitting or detecting signals, (ii) of classifying records of phenomena such as brain waves, (iii) of studying radio propagation

phenomena, and (iv) of determining characteristics of control systems. The theory of statistical spectral analysis is too extensive to be reviewed here. For surveys of this theory, see Bartlett [3], Hannan [17], Blackman and Tukey [4], Jenkins [20], Parzen [41], Rosenblatt [45], and Tukey [53].

A comprehensive survey of the results available on topics (i) and (ii) has been given recently by Hannan [17]. Other comprehensive reviews are given by Bartlett [3], Moran [33], and Wold [57], Chap. 11; in these reviews one may find references to the work of M. S. Bartlett, H. E. Daniels, J. Durbin, E. J. Hannan, G. M. Jenkins, M. H. Quenouille, A. M. Walker, G. S. Watson, and P. Whittle. In Section 8, I have attempted to give an introduction to some of the large sample results available on topics (i) and (ii).

(II) *Prediction theory and the Hilbert space spanned by a time series.* A basic problem in time series analysis is that of *minimum mean square error linear prediction*. Let  $Z$  be an unobserved random variable with finite second moment. Let  $\{X(t), t \in T\}$  be an observed time series. One seeks that random variable, linear in the observations, whose mean square distance from  $Z$  is smallest. In other words, if one desires to predict the value of  $Z$ , on the basis of having observed the values of the time series  $\{X(t), t \in T\}$ , one method might be to take that linear functional in the observations, denoted by  $E^*[Z | X(t), t \in T]$ , whose mean square error as a predictor is least. (The symbol  $E^*$  is used to denote a predictor because in the case of jointly normally distributed random variables, the best linear predictor  $E^*[Z | X(t), t \in T]$  coincides with the conditional expectation  $E[Z | X(t), t \in T]$ ; for an elementary discussion of this fact, see Parzen ([38], p. 387). Indeed, it should be noted that in any event the conditional expectation  $E[Z | X(t), t \in T]$  can be defined as the minimum mean square error *non-linear* predictor.)

The prediction problem has provided a framework in terms of which many problems of statistical communication theory have come to be formulated. The pioneering work on prediction theory was done by Wiener [56a] and Kolmogorov [25] who were concerned with a stationary time series which had been observed over a semi-infinite interval of time. They sought predictors which had minimum mean square over all possible linear predictors. Wiener showed how the solution of the prediction problem could be reduced to the solution of the so-called Wiener-Hopf integral equation, and gave a method (spectral factorization) for the solution of the integral equation. Simplified methods of solution of this equation in the practically important special case of rational spectral density functions were given by Zadeh and Ragazzini [61] and Bode and Shannon [5]. Zadeh and Ragazzini [62] also treated the problem of regression analysis of time series with stationary fluctuation function, by reducing the problem to one involving the solution of a Wiener-Hopf equation. There then developed an extensive literature, seeking to treat prediction and smoothing problems involving a finite time of observation and non-stationary time series. The methods employed were either to reduce the problem to the solution of a suitable integral equation (generalization of the Wiener-Hopf equation) or to employ expansions (in a series of suitable eigen functions) of the time series involved.



As a result of these developments, prediction theory has turned out to provide a theory of the structure of time series and to provide mathematical tools for the solutions of other problems besides the prediction problem, especially regression problems. In Section 3, it is shown how the prediction problem leads naturally to the introduction of the important notion of the *Hilbert space spanned by a time series* which plays a central role in modern time series analysis. In Sections 4 and 6, I describe an approach to prediction and regression problems (in terms of reproducing kernel Hilbert spaces) which may be called coordinate free, and which by the introduction of suitable coordinate systems contains previous approaches as special cases.

The approach I take seems to me to be a rigorous version of an approach that is being developed in the Soviet Union by V. S. Pugachev [44]. Pugachev has in recent years advanced a point of view, which he calls the method of canonic representations of random functions, for which in one of his articles [43] he makes the following claim. "The results of this article, together with the results of [previous] papers, permits us to state that the method of canonic representations of random functions is the foundation of the modern statistical theory of optimum systems." It is my feeling that reproducing kernel Hilbert spaces provide a more powerful and more elegant means of achieving in a unified manner the results which Pugachev has sought to unify by the method of canonic representations.

(III) *Regression analysis of time series with known covariance function:* Let the observed time series be of the form of (1.1) with unknown mean value function  $m(t)$  and known covariance function  $K(s, t)$ . Various methods of forming estimates of  $m(t)$  are available. The most important methods are classical least squares estimation and minimum variance linear unbiased estimation. In the case of normally distributed observations, one has in addition the methods of maximum likelihood estimation and minimum variance unbiased estimation. In Sections 6 and 7, it is shown how Hilbert space techniques may be used to form explicit expressions for these estimates in terms of certain so-called reproducing kernel inner products.

There are, of course, large numbers of important problem areas of time series analysis which have not been mentioned in the foregoing such as (i) the problem of the distribution of zero-crossings and extrema of a time series (see references, see Longuet-Higgins [30] and Slepian [50]), (ii) the problem of the asymptotic efficiency of various classes of estimates of regression coefficients (see Grenander and Rosenblatt [13], Hebbe [19], and Striebel [52]), (iii) the use of filters to eliminate or extract trend or other components of a time series, and (iv) the distribution of various functionals of a time series, such as quadratic forms.

Further, the statistical analysis of multiple time series is not discussed. The relations that exist between different time series is on the whole a problem of greater interest than the relations that exist within a single time series. The results which exist under categories (I), (II), and (III) for univariate time series can be formally extended to multiple time series. However, many new problems arise which have not been thoroughly investigated.

A word should be said about the references given at the end of this paper. I have given a representative list rather than a complete list. Fortunately, a complete list of references will soon be available. The International Statistical Institute is compiling a bibliography on *Time Series and Stochastic Processes* which is to list and classify books and papers published, in the years 1900-1959, on both theory and applications. A bibliography (Parzen [42a]) of American publications has been compiled at Stanford for inclusion in the I.S.I. bibliography; a limited number of copies of this bibliography are available, and may be obtained by writing to the author. A bibliography is also given by Deming [8].

## 2. Stationary time series. A discrete parameter time series

$$\{X(t), t = 0, \pm 1, \dots\}$$

or a continuous parameter time series  $\{X(t), -\infty < t < \infty\}$  is said to be (weakly or wide-sense) stationary if the product moment

$$E[X(s)X(s+t)] = R(t)$$

is a function only of  $t$ . One calls  $R(\cdot)$  the covariance function of the stationary time series.

It was shown by Khintchine [24] in 1933 that in the continuous parameter case there exists a non-decreasing bounded function  $F(\omega)$ , defined for  $-\infty < \omega < \infty$ , such that

$$(2.1) \quad R(t) = \int_{-\infty}^{\infty} e^{it\omega} dF(\omega), \quad -\infty < t < \infty,$$

if it is assumed that  $R(\cdot)$  is continuous at  $t = 0$ . Wold [58] in 1938 showed that in the discrete parameter case there exists a non-decreasing bounded function  $F(\omega)$ , defined for  $-\pi \leq \omega \leq \pi$ , such that

$$(2.2) \quad R(t) = \int_{-\pi}^{\pi} e^{it\omega} dF(\omega), \quad t = 0, \pm 1, \dots$$

The function  $F(\omega)$  is called the *spectral distribution function* of the time series. Like a probability distribution function,  $F(\omega)$  can be uniquely written as the sum,

$$F(\omega) = F_d(\omega) + F_{sc}(\omega) + F_{ac}(\omega),$$

of three distribution functions with the following properties. The function  $F_{ac}(\omega)$  is absolutely continuous and is the integral of a non-negative function  $f(\omega)$  called the *spectral density function* of the time series. The function  $F_d(\omega)$  is a purely discontinuous (or discrete or step) function:

$$F_d(\omega) = \sum_{\omega_j \leq \omega} \Delta F(\omega_j)$$

where  $\{\omega_j\}$  are the discontinuity points of  $F(\omega)$ , and  $\Delta F(\omega) = F(\omega + 0) - F(\omega - 0)$ . Finally,  $F_{sc}(\omega)$  is a singular continuous function.

It is usually assumed that physically observed time series have a spectral distribution function of the following form:

$$F(\omega) = \sum_{\substack{\omega' \text{ such} \\ \text{that } \Delta F(\omega') > 0 \\ \text{and } \omega' \leq \omega}} \Delta F(\omega') + \int_{-\infty}^{\omega} f(\omega') d\omega'$$

where (i) the spectral density function  $f(\omega)$  has the property that it is an integrable non-negative function which is continuous except at a finite number of points where it has finite left-hand and right-hand limits, and (ii) the set of frequencies at which the spectral jump function (or spectral mass function)  $\Delta F(\omega)$  is positive contains at most countably infinite many points distributed on the real line in such a way that in any finite interval there are only a finite number of points of positive spectral mass. If these conditions are satisfied, we say that the time series has a *mixed spectrum*. If the spectral density function vanishes for all  $\omega$ , we say that the time series has a *discrete spectrum*. If the spectral jump function  $\Delta F(\omega)$  vanishes for all  $\omega$ , we say that the time series has a *continuous spectrum*.

In terms of the spectral distribution function one can characterize various representations (or models) for a stationary time series  $X(t)$ . For example, it may be shown that a discrete parameter time series with a mixed spectrum whose spectral density function satisfies the condition

$$\int_{-\pi}^{\pi} \log f(\omega) d\omega > -\infty$$

may be written

$$(2.3) \quad X(t) = \sum_r A_r e^{i\omega_r t} + \sum_{v=0}^{\infty} c_v \eta(t-v)$$

for suitable sequences of frequencies  $\{\omega_r\}$ , constants  $\{c_r\}$ , and uncorrelated random variables  $\{A_r\}$  and  $\{\eta_r\}$ . In view of (2.3) one sees that the scheme of hidden periodicities and the scheme of moving averages may be viewed as a special kind of stationary process. Similarly, it may be shown that an autoregressive scheme (where the  $\eta_t$  are uncorrelated rather than independent) corresponds to a stationary time series whose distribution function is absolutely continuous and whose spectral density function is of the form

$$f(\omega) = \left[ 2\pi \left| \sum_{k=0}^m b_k e^{ik\omega} \right|^2 \right]^{-1}$$

for suitable constants  $b_0, \dots, b_m$ . To prove these assertions, one uses the Hilbert space representation theory described in Section 4.

**3. The problem of minimum mean square error linear prediction.** In order to show existence and uniqueness, and to obtain conditions characterizing, the best

linear predictor, we need to introduce the notion of a Hilbert space. (For a discussion of Hilbert space theory see any suitable text, such as Halmos [16].)

DEFINITION 3A. By an abstract Hilbert space is meant a set  $H$  whose members  $u, v, \dots$  are usually called vectors or points which possesses the following properties.

(I)  $H$  is a linear space [that is, for any vectors  $u$  and  $v$  in  $H$ , and real number  $a$ , there exist vectors, denoted by  $u + v$  and  $au$  respectively, which satisfy the usual algebraic properties of addition and multiplication; also there exists a zero vector  $0$  with the usual properties under addition].

(II)  $H$  is an inner product space [that is, to every pair of points  $u$  and  $v$  in  $H$  there corresponds a real number, written  $(u, v)$  and called the inner product of  $u$  and  $v$ , possessing the following properties: for all points  $u, v$ , and  $w$  in  $H$ , and every real number  $a$ ,

- (i)  $(au, v) = a(u, v)$
- (ii)  $(u + v, w) = (u, w) + (v, w)$
- (iii)  $(v, u) = (u, v)$
- (iv)  $(u, u) > 0$  if and only if  $u \neq 0$ ].

(III)  $H$  is a complete metric space under the norm  $\|u\| = (u, u)^{1/2}$  [that is, if  $\{u_n\}$  is a sequence of points such that  $\|u_m - u_n\| \rightarrow 0$  as  $m, n \rightarrow \infty$  then there is a vector  $u$  in  $H$  such that  $\|u_n - u\|^2 \rightarrow 0$  as  $n \rightarrow \infty$ ].

In order to define the notion of the Hilbert space spanned by a time series, we first define the notion of the Hilbert space spanned by a family of vectors.

DEFINITION 3B. Let  $T$  be an index set, and let  $\{u(t), t \in T\}$  be a family of members of a Hilbert space  $H$ . The linear manifold spanned by the family  $\{u(t), t \in T\}$ , denoted  $L(u(t), t \in T)$ , is defined to be the set, consisting of all vectors  $u$  in  $H$  which may be represented in the form  $u = \sum_{i=1}^n c_i u(t_i)$  for some integer  $n$ , some constants  $c_1, \dots, c_n$ , and some points  $t_1, \dots, t_n$  in  $T$ . The Hilbert space spanned by the family  $\{u(t), t \in T\}$ , denoted  $V(u(t), t \in T)$  [or  $L_2(u(t), t \in T)$  if  $H$  is the space of square integrable functions on some measure space], is defined to be the set of vectors which either belong to the linear manifold  $L(u(t), t \in T)$  or may be represented as a limit of vectors in  $L(u(t), t \in T)$ . If  $V(u(t), t \in T)$  coincides with  $H$ , we say that  $\{u(t), t \in T\}$  spans  $H$ .

DEFINITION 3C. The Hilbert space spanned by a time series  $\{X(t), t \in T\}$ , denoted by  $L_2(X(t), t \in T)$ , is defined to consist of all random variables  $U$  which are either finite linear combinations of the random variables  $\{X(t), t \in T\}$  or are limits of such finite linear combinations in the norm corresponding to the inner product defined on the space  $L_2$  of square integrable random variables by

$$(3.1) \quad (U, V) = E[UV].$$

In words,  $L_2(X(t), t \in T)$  consists of all linear functionals in the time series.

We next state without proof the projection theorem for an abstract Hilbert space.

PROJECTION THEOREM. Let  $H$  be an abstract Hilbert space, let  $M$  be a Hilbert subspace of  $H$ , let  $v$  be a vector in  $H$ , and let  $v^*$  be a vector in  $M$ . A necessary and



sufficient condition that  $v^*$  is the unique vector in  $M$  satisfying

$$(3.2) \quad \|v^* - v\| = \min_{u \in M} \|u - v\|$$

is that

$$(3.3) \quad (v^*, u) = (v, u) \quad \text{for every } u \text{ in } M.$$

The vector  $v^*$  satisfying (3.2) is called the projection of  $v$  onto  $M$ , and will here be written  $E^*[v | M]$ .

In the case that  $M$  is the Hilbert space spanned by a family of vectors  $\{x(t), t \in T\}$  in  $H$ , we write  $E^*[v | x(t), t \in T]$  to denote the projection of  $v$  onto  $M$ . In this case, a necessary and sufficient condition that  $v^*$  satisfy (3.3) is that

$$(3.4) \quad (v^*, x(s)) = (v, x(s)) \quad \text{for every } s \text{ in } T.$$

We are now in a position to solve the problem of obtaining an explicit expression for the minimum mean square error linear prediction  $E^*[Z | X(t), t \in T]$ . From (3.4), with  $H$  equal to the Hilbert space  $L_2$  of all square integrable random variables, and  $v = Z$ , it follows that the optimum linear predictor is the unique random variable in  $L_2(X(t), t \in T)$  satisfying, for all  $s$  in  $T$ ,

$$(3.5) \quad E[E^*[Z | X(t), t \in T]X(s)] = E[ZX(s)].$$

Equation (3.5) may look more familiar if we consider the special case of an interval  $T = \{t: a \leq t \leq b\}$ . If one writes heuristically

$$(3.6) \quad \int_a^b X(t)w(t) dt$$

to represent a random variable in  $L_2(X(t), t \in T)$ , then (3.5) states that the weighting function  $w^*(t)$  of the best linear predictor

$$(3.7) \quad E^*[Z | X(t), t \in T] = \int_a^b w^*(t)X(t) dt,$$

must satisfy the generalized Wiener-Hopf equation

$$(3.8) \quad \int_a^b w^*(t)K(s, t) dt = \rho_Z(s), \quad a \leq s \leq b$$

where we define

$$(3.9) \quad K(s, t) = E[X(s)X(t)]$$

$$(3.10) \quad \rho_Z(t) = E[ZX(t)].$$

There is an extensive literature concerning the solution of the integral equation in (3.8); see [39] for references. In my opinion, however, this literature is concerned with an unnecessarily hard problem, as well as one in which the very formulation of the problem makes it difficult to be rigorous. The integral equation in (3.8) possesses a solution only if one interprets  $w^*(t)$  as a generalized

function which includes terms which are Dirac delta functions and derivatives of delta functions.

It seems to me that a simple reinterpretation of (3.8) avoids all these difficulties. Let us not regard (3.8) as an integral equation for the weighting function  $w^*(t)$ . Rather, let us compare (3.7) and (3.8). These equations say that if one can find a representation for the function  $\rho_z(s)$  in terms of linear operations on the functions  $\{K(s, t), t \in T\}$ , then the minimum mean square error linear predictor  $E^*[Z | X(t), t \in T]$  can be written in terms of the corresponding linear operations on the time series  $\{X(t), t \in T\}$ . It should be emphasized that the most important linear operations are integration and differentiation. Consequently, the problem of finding the best linear predictor is not one of solving an integral equation, but is one of hunting for a linear representation of  $\rho_z(t)$  in terms of the covariance kernel  $K(s, t)$ . A general method of finding such representations will be discussed in Sections 4 and 5.

We illustrate the ideas involved by considering a simple example.

EXAMPLE 3A. Consider a stationary time series  $X(t)$ , with covariance kernel

$$(3.11) \quad K(s, t) = Ce^{-\beta|t-s|},$$

which one has observed over a finite interval of time,  $a \leq t \leq b$ . Suppose that one desires to predict  $X(b+c)$ , for  $c > 0$ . Now, for  $a \leq t \leq b$ ,

$$(3.12) \quad \rho(t) = E[X(t)X(b+c)] = Ce^{-\beta(b+c-t)} = e^{-\beta c}K(b, t).$$

In view of (3.12), by the interpretation of (3.7) and (3.8) just stated, it follows that

$$(3.13) \quad E^*[X(b+c) | X(t), a \leq t \leq b] = e^{-\beta c}X(b).$$

**4. Hilbert space representations of time series.** In the decade of the 1940's, probabilists began to employ Hilbert space methods to clarify the structure of time series (see [21] and [27]). Among the fundamental theorems proved in this period were the spectral representation theorem for stationary time series, and the Karhunen-Loève representation for random functions of second order on a finite interval. Various workers (especially Grenander [12]) have made use of these representation theorems in treating problems of statistical inference on time series. A representation theorem which does not seem to have found any application is one due to Loève ([27], p. 338) which shows that there is a very intimate connection between time series (random functions of second order) and reproducing kernel Hilbert spaces. It turns out, in my opinion, that reproducing kernel Hilbert spaces are the natural setting in which to solve problems of statistical inference on time series. In this section we define the notion of a Hilbert space representation of a time series and show how this notion may be used to explicitly solve the prediction problem.

The definition we give of the notion of a Hilbert space representation of a time series is based on the following theorem (for proof, see Parzen [37] or [40]).

**BASIC CONGRUENCE THEOREM.** Let  $H_1$  and  $H_2$  be two abstract Hilbert spaces.

Denote the inner product between two vectors  $u_1$  and  $u_2$  in  $H_1$  by  $(u_1, u_2)_1$ . Similarly, denote the inner product between two vectors  $v_1$  and  $v_2$  in  $H_2$  by  $(v_1, v_2)_2$ . Let  $T$  be an index set. Let  $\{u(t), t \in T\}$  be a family of vectors which span  $H_1$ . Similarly, let  $\{v(t), t \in T\}$  be a family of vectors which span  $H_2$ . Suppose that, for every  $s$  and  $t$  in  $T$ ,

$$(4.1) \quad (u(s), u(t))_1 = (v(s), v(t))_2.$$

Then there exists a congruence (a one-one inner product preserving linear mapping)  $\psi$  from  $H_1$  onto  $H_2$  which has the property that

$$(4.2) \quad \psi(u(t)) = v(t), \quad t \text{ in } T.$$

DEFINITION 4A. A family of vectors  $\{f(t), t \in T\}$  in a Hilbert space  $H$  is said to be a representation of a time series  $\{X(t), t \in T\}$  if, for every  $s$  and  $t$  in  $T$ ,

$$(4.3) \quad (f(s), f(t))_H = K(s, t) = E[X(s)X(t)].$$

Then there is a congruence (a one-one inner product preserving linear mapping)  $\psi$  from  $V(f(t), t \in T)$  onto  $L_2(X(t), t \in T)$  satisfying

$$(4.4) \quad \psi(f(t)) = X(t)$$

and every random variable  $U$  in  $L_2(X(t), t \in T)$  may be written

$$(4.5) \quad U = \psi(g)$$

for some unique vector  $g$  in  $V(f(t), t \in T)$ .

We next show that the representation of a time series as a stochastic integral is best viewed as a Hilbert space representation.

DEFINITION 4B. We call  $(Q, \mathbf{B}, \mu)$  a measure space if  $Q$  is a set,  $\mathbf{B}$  is a  $\sigma$ -field of subsets of  $Q$ , and  $\mu$  is a measure on the measurable space  $(Q, \mathbf{B})$ . We denote by  $L_2(Q, \mathbf{B}, \mu)$  the Hilbert space of all  $B$ -measurable real valued functions defined on  $Q$  satisfying

$$(4.6) \quad (f, f)_\mu = \int_Q f^2 d\mu < \infty.$$

DEFINITION 4C. Let  $(Q, \mathbf{B}, \mu)$  be a measure space, and, for every  $B$  in  $\mathbf{B}$ , let  $Z(B)$  be a random variable. The family of random variables  $\{Z(B), B \in \mathbf{B}\}$  is called an *orthogonal random set function* with covariance kernel  $\mu$  if, for any two sets  $B_1$  and  $B_2$  in  $\mathbf{B}$ ,

$$(4.7) \quad E[Z(B_1)Z(B_2)] = \mu(B_1 B_2),$$

where, as usual,  $B_1 B_2$  denotes the intersection of  $B_1$  and  $B_2$ .

The Hilbert space  $L_2(Z(B), B \in \mathbf{B})$  of random variables spanned by an orthogonal random set function may be defined, as was the Hilbert space spanned by a time series, to be the smallest Hilbert subspace of the Hilbert space of all square integrable random variables containing all random variables  $U$  of the form  $U = \sum_{i=1}^n c_i Z(B_i)$  for some integer  $n$ , subfamily  $\{B_1, \dots, B_n\} \subset \mathbf{B}$ , and real

constants  $c_1, \dots, c_n$ . On the other hand,  $L_2(Q, \mathbf{B}, \mu)$  may be described as the Hilbert space spanned under the norm (4.6) by the family of indicator functions  $(I_B, B \in \mathbf{B})$ , where the indicator function  $I_B$  of  $B$  is defined by  $I_B(q) = 1$  or 0 according as  $q \in B$  or  $q \notin B$ . Now for any  $B_1$  and  $B_2$  in  $\mathbf{B}$ ,

$$(4.8) \quad (I_{B_1}, I_{B_2})_\mu = \mu(B_1 B_2) = E[Z(B_1)Z(B_2)].$$

Therefore, by the Basic Congruence Theorem, there is a congruence  $\psi$  from  $L_2(Q, \mathbf{B}, \mu)$  onto  $L_2(Z(B), B \in \mathbf{B})$  such that for any  $B \in \mathbf{B}$ ,

$$(4.9) \quad \psi(I_B) = Z(B).$$

This fact justifies the following definition of the stochastic integral.

DEFINITION 4D. Let  $(Q, \mathbf{B}, \mu)$  be a measure space and let  $\{Z(B), B \in \mathbf{B}\}$  be an orthogonal random set function with covariance kernel  $\mu$ . For any function  $f$  in  $L_2(Q, \mathbf{B}, \mu)$  one defines the stochastic integral of  $f$  with respect to  $\{Z(B), B \in \mathbf{B}\}$ , denoted  $\int_Q f dZ$ , by

$$(4.10) \quad \int_Q f dZ = \psi(f),$$

where  $\psi$  is the congruence from  $L_2(Q, \mathbf{B}, \mu)$  onto  $L_2(Z(B), B \in \mathbf{B})$  determined by (4.9).

We are now in a position to state our version of Karhunen's theorem (see [13], p. 29).

THEOREM 4A. Let  $\{X(t), t \in T\}$  be a time series with covariance kernel  $K$ . Let  $\{f(t), t \in T\}$  be a family of functions in a space  $L_2(Q, \mathbf{B}, \mu)$ , such that for all  $s, t$  in  $T$

$$(4.11) \quad K(s, t) = \int_Q f(s)f(t) d\mu.$$

Then  $\{f(t), t \in T\}$  is a representation for  $\{X(t), t \in T\}$ .

If, further,  $\{f(t), t \in T\}$  spans  $L_2(Q, \mathbf{B}, \mu)$ , then there is an orthogonal random set function  $\{Z(B), B \in \mathbf{B}\}$  with covariance kernel  $\mu$  such that

$$(4.12) \quad X(t) = \int_Q f(t) dZ, \quad t \in T,$$

and every random variable  $U$  in  $L_2(X(t), t \in T)$  may be represented

$$(4.13) \quad U = \int_Q g dZ$$

for some unique function  $g$  in  $L_2(Q, \mathbf{B}, \mu)$ .

PROOF. Let  $\psi$  be the congruence from  $L_2(f(t), t \in T)$  onto  $L_2(X(t), t \in T)$  satisfying (4.4). If  $\{f(t), t \in T\}$  spans  $L_2(Q, \mathbf{B}, \mu)$ , define, for  $B \in \mathbf{B}$ ,  $Z(B) = \psi(I_B)$ . It is immediate that  $\{Z(B), B \in \mathbf{B}\}$  is an orthogonal random set function with covariance kernel  $\mu$ . By the definition of the stochastic integral, (4.12) is merely another way of writing the fact that  $X(t) = \psi(f(t))$ .

Theorem 4A, together with (2.2) and (2.1), yields the following fundamental result.



**SPECTRAL REPRESENTATION THEOREM FOR STATIONARY TIME SERIES.** A discrete parameter time series  $\{X(t), t = 0, \pm 1, \dots\}$  is weakly stationary if and only if for some Lebesgue-Stieltjes measures  $\mu$  on the interval  $Q = \{\lambda: -\pi \leq \lambda \leq \pi\}$  the complex exponentials  $\{e^{i\lambda t}, t = 0, \pm 1, \dots\}$  form a representation for the time series in  $L_2(Q, \mathbf{B}, \mu)$  where  $\mathbf{B}$  is the  $\sigma$ -field of Borel subsets of  $Q$ . Then there exists an orthogonal random set function  $\{Z(B), B \in \mathbf{B}\}$  such that

$$(4.14) \quad X(t) = \int_{-\pi}^{\pi} e^{i\lambda t} Z(d\lambda), t = 0, \pm 1, \dots$$

A similar theorem holds for continuous parameter time series with

$$Q = \{\lambda: -\infty < \lambda < \infty\}.$$

The representation of a time series as an integral with respect to an orthogonal random set function is not a natural representation, since one may choose such representations of a time series in a multitude of ways. Indeed, if  $(Q, \mathbf{B}, \mu)$  is a measure space such that  $L_2(X(t), t \in T)$  and  $Q_2(Q, \mathbf{B}, \mu)$  have the same dimension, there are many families  $\{f(t), t \in T\}$  of functions in  $L_2(Q, \mathbf{B}, \mu)$  which are a representation for  $\{X(t), t \in T\}$ . What one desires is a family  $\{f(t), t \in T\}$  of familiar functions [such as the family of complex exponentials  $e^{i\lambda t}$ , which are a representation in a suitable space  $L_2(Q, \mathbf{B}, \mu)$  for a stationary time series]. I believe there is a natural representation in terms of which to solve problems of statistical inference on time series, namely the representation of a time series with covariance kernel  $K$  by the functions  $\{K(\cdot, t), t \in T\}$  in the reproducing kernel Hilbert space  $H(K)$ .

**DEFINITION 4E.** A Hilbert space  $H$  is said to be a reproducing kernel Hilbert space, with reproducing kernel  $K$ , if the members of  $H$  are functions on some set  $T$ , and if there is a kernel  $K$  on  $T \otimes T$  having the following two properties; for every  $t$  in  $T$  (where  $K(\cdot, t)$  is the function defined on  $T$ , with value at  $s$  in  $T$  equal to  $K(s, t)$ ):

$$(4.15) \quad K(\cdot, t) \in H$$

$$(4.16) \quad (g, K(\cdot, t))_H = g(t)$$

for every  $g$  in  $H$ .

Intuitively, a reproducing kernel Hilbert space is a Hilbert space which contains a function playing the role of the Dirac delta function  $\delta(t)$ . It should be recalled that, for square integrable functions  $f(\cdot)$ ,

$$\int_{-\infty}^{\infty} f(s) \delta(s - t) ds = f(t).$$

Consequently, the kernel  $K(s, t) = \delta(s - t)$  satisfies (4.16). However it does not satisfy (4.15), and therefore is not truly a reproducing kernel.

**THEOREM 4B** (Moore-Aronszajn-Loève [1], [27]). The covariance kernel  $K$  of a time series generates a unique Hilbert space, which we denote by  $H(K)$ , of which  $K$  is the reproducing kernel.

Since  $K(s, t) = (K(\cdot, s), K(\cdot, t))_{H(K)} = E[X(s)X(t)]$  we immediately obtain the following important theorem.

**THEOREM 4C.** Let  $\{X(t), t \in T\}$  be a time series with covariance kernel  $K$ . Then the family  $\{K(\cdot, t), t \in T\}$  of functions in  $H(K)$  is a representation for  $\{X(t), t \in T\}$ . Given a function  $g$  in  $H(K)$ , we denote by  $(X, g)_K$  or  $(g, X)_K$  the random variable  $U$  in  $L_2(X(t), t \in T)$  which corresponds to  $g$  under the congruence which maps  $K(\cdot, t)$  into  $X(t)$ . We then have the following formal relations: for every  $t$  in  $T$ , and  $g, h$  in  $H(K)$ ,

$$(4.17) \quad \begin{aligned} (X, K(\cdot, t))_K &= X(t) \\ E[(X, h)_K(X, g)_K] &= (h, g)_K \end{aligned}$$

where we hereafter write  $(h, g)_K$  for  $(h, g)_{H(K)}$ .

The next theorem shows the relationship between the reproducing kernel Hilbert space representation of a time series, and the representation of a time series by an orthogonal decomposition of the form of (4.12).

**THEOREM 4D.** Let  $K$  be a covariance kernel. If there exist a measure space  $(Q, \mathbf{B}, \mu)$ , and a family of functions  $\{f(t), t \in T\}$  in  $L_2(Q, \mathbf{B}, \mu)$  such that (4.11) holds, then the reproducing kernel Hilbert space  $H(K)$  corresponding to the covariance kernel  $K$  may be described as follows:  $H(K)$  consists of all functions  $g$ , defined on  $T$ , which may be represented in the form

$$(4.18) \quad g(t) = \int_Q g^* f(t) d\mu$$

for some (necessarily unique) function  $g^*$  in the Hilbert subspace  $L_2(f(t), t \in T)$  of  $L_2(Q, \mathbf{B}, \mu)$  spanned by the family of functions  $\{f(t), t \in T\}$ , with norm given by

$$(4.19) \quad \|g\|^2 = \int_Q |g^*|^2 d\mu.$$

If  $\{f(t), t \in T\}$  spans  $L_2(Q, \mathbf{B}, \mu)$ , so that  $X(t)$  has an orthogonal decomposition (4.12), then we may write

$$(4.20) \quad (X, g)_K = \int_Q g^* dZ.$$

**PROOF.** Verify that the set  $H$  of functions of the form of (4.18), with norm given by (4.19), is a Hilbert space satisfying (4.15) and (4.16).

**THEOREM 4E.** (General solution of the prediction problem.) Let  $\{X(t), t \in T\}$ , be a time series with covariance kernel  $K(s, t)$ , and let  $H(K)$  be the corresponding reproducing kernel Hilbert space. Between  $L_2(X(t), t \in T)$  and  $H(K)$  there exists a one-one inner product preserving linear mapping under which  $X(t)$  and  $K(\cdot, t)$  are mapped into one another. Denote by  $(h, X)_K$  the random variable in  $L_2(X(t), t \in T)$  which corresponds under the mapping to the function  $h(\cdot)$  in  $H(K)$ . Then the general solution to the prediction problem may be written as follows. If  $Z$  is a random variable with finite second moment, and if

$$(4.21) \quad \rho_Z(t) = E[ZX(t)],$$

then

$$(4.22) \quad E^*[Z | X(t), t \in T] = (\rho_Z, X)_K$$

with mean square error of prediction given by

$$(4.23) \quad E[|Z - E^*[Z | X(t), t \in T]|^2] = E|Z|^2 - (\rho_Z, \rho_Z)_K.$$

Theorem 4E represents a coordinate free solution of the prediction problem. The usual methods of explicitly writing optimum predictors, using either eigenfunction expansions, Green's functions (impulse response function), or (power) spectral density functions, are merely methods of writing down the reproducing kernel inner product corresponding to the covariance kernel  $K(s, t)$  of the observed time series.

The validity of Theorem 4E follows immediately from the definition of the concepts involved. However, it may be instructive to give a proof of the theorem, using the following properties of the mapping  $(h, X)_K$ . For any functions  $g$  and  $h$  in  $H(K)$  and random variables  $Z$  with finite second moment it holds that

$$(4.24) \quad E[(h, X)_K(g, X)_K] = (h, g)_K$$

$$(4.25) \quad E[Z(h, X)_K] = (\rho_Z, h)_K,$$

in which  $\rho_Z(t) = E[ZX(t)]$ . Now a random variable in  $L_2(X(t), t \in T)$  may be written  $(h, X)_K$  for some  $h$  in  $H(K)$ . Consequently the mean square error between any linear functional  $(h, X)_K$  and  $Z$  may be written

$$(4.26) \quad \begin{aligned} E[|(h, X)_K - Z|^2] &= E[(h, X)_K^2] + E[Z^2] - 2E[Z(h, X)_K] \\ &= E[Z^2] + (h, h)_K - 2(\rho_Z, h)_K \\ &= E[Z^2] - (\rho_Z, \rho_Z)_K + (h - \rho_Z, h - \rho_Z)_K. \end{aligned}$$

From (4.26) it is immediate that  $(\rho_Z, X)_K$  is the minimum mean square error linear predictor of  $Z$ , with mean square prediction error equal to  $E[Z^2] - (\rho_Z, \rho_Z)_K$ . The proof of Theorem 4E is complete.

**5. Examples of reproducing kernel Hilbert space representations.** In this section we give the reproducing kernel Hilbert space representation of a time series  $\{X(t), t \in T\}$  under a variety of standard assumptions.

**EXAMPLE 5A.** Suppose  $T = \{1, 2, \dots, N\}$  for some positive integer  $N$ , and that the covariance kernel  $K$  is given by a symmetric positive definite matrix  $\{K_{ij}\}$  with inverse  $\{K^{ij}\}$ . The corresponding reproducing kernel space  $H(K)$  consists of all  $N$ -dimensional vectors  $f = (f_1, \dots, f_N)$  with inner product

$$(5.1) \quad (f, g)_K = \sum_{s,t=1}^N f_s K^{st} g_t.$$

To prove (5.1) one need only verify that the reproducing property holds: for  $u = 1, \dots, N$ ,

$$(f, K_{\cdot u})_K = \sum_{s,t=1}^N f_s K^{st} K_{tu} = \sum_{s=1}^N f_s \delta(s, u) = f_u.$$

The inner product may also be written as a ratio of determinants:

$$(5.2) \quad (f, g)_K = - \frac{\begin{vmatrix} K_{11} & \cdots & K_{1N} & f_1 \\ \vdots & \cdots & \vdots & \vdots \\ K_{N1} & \cdots & K_{NN} & f_N \\ g_1 & \cdots & g_N & 0 \end{vmatrix}}{\begin{vmatrix} K_{11} & \cdots & K_{1N} \\ \vdots & \cdots & \vdots \\ K_{N1} & \cdots & K_{NN} \end{vmatrix}}.$$

To prove (5.2) one again need only verify the reproducing property. In the case in which the covariance matrix  $K$  is singular, one may define the corresponding reproducing kernel inner product in terms of the *pseudo-inverse* of the matrix  $K$  (see Greville [15] for a discussion of the notion of pseudo-inverse).

Although (5.1) provides a formula for the reproducing kernel inner product in terms of the inverse of the covariance matrix, it is to be emphasized that one need not necessarily invert the covariance matrix in order to find the reproducing kernel inner product. The point of introducing the reproducing kernel inner product is that the inversion of the covariance matrix is usually an intractable problem, and one should look instead to evaluate that for which one would use the inverse  $K^{-1}$ ; namely, the evaluation of inner products in the reproducing kernel space. Various iterative methods of evaluating these inner products can be given (see [39] or [40]). This observation is undoubtedly not as important in the case of discrete parameter time series as it is in the case of multiple time series and continuous parameter time series.

**EXAMPLE 5B. Autoregressive schemes (discrete parameter).** A discrete parameter weakly stationary time series  $X(t)$  is said to satisfy an autoregressive scheme of order  $m$  if  $X(t)$  is the solution of the stochastic difference equation

$$(5.3) \quad L_t X(t) = \sum_{k=0}^m a_k X(t-k) = \eta(t)$$

where  $a_0, \dots, a_m$  are given constants, and  $\{\eta(t)\}$  is an orthonormal sequence of random variables. We now show that given observations  $\{X(t), t = 1, 2, \dots, N\}$  the reproducing kernel Hilbert space  $H(K)$  corresponding to the covariance kernel  $K$  of the observations consists of all  $N$ -vectors  $f = (f(1), \dots, f(N))$  with inner product given by

$$(5.4) \quad (f, g)_K = \sum_{t=m+1}^N \{L_t f(t)\} \{L_t g(t)\} + \sum_{j,k=1}^m d_{jk} f(j) g(k)$$

where the matrix  $D = \{d_{jk}\}$  has an inverse  $D^{-1} = \{d^{jk}\}$  with general term

$$(5.5) \quad d^{jk} = K(j-k) = E[X(j)X(k)].$$

In the case that  $N \geq 2m$ , an explicit expression for  $d_{jk}$  is given by

$$(5.6) \quad d_{jk} = \sum_{u=1}^{\min(j,k)} \{a_{j-u} a_{k-u} - a_{u+m-j} a_{u+m-k}\}.$$

In particular for a first order autoregressive scheme and  $N \geq 2$

$$(5.7) \quad \begin{aligned} (f, g)_K &= (a_0^2 - a_1^2)f(1)g(1) \\ &+ \sum_{t=2}^N \{a_0f(t) + a_1f(t-1)\}\{a_0g(t) + a_1g(t-1)\}. \end{aligned}$$

For a second order autoregressive scheme and  $N \geq 4$

$$(5.8) \quad \begin{aligned} (f, g)_K &= (a_0^2 - a_2^2)\{f(1)g(1) + f(2)g(2)\} \\ &+ (a_0a_1 - a_1a_2)\{f(1)g(2) + f(2)g(1)\} \\ &+ \sum_{t=3}^N \{a_0f(t) + a_1f(t-1) + a_2f(t-2)\} \\ &\cdot \{a_0g(t) + a_1g(t-1) + a_2g(t-2)\}. \end{aligned}$$

One can give a purely algebraic proof of (5.4). However a simpler proof can be given if one uses certain facts from probability theory. Let us suppose that  $X(1), \dots, X(N)$  are jointly normally distributed random variables with covariance matrix  $K_N = \{K_{s,t}\}$  with inverse matrix  $K_N^{-1} = \{K^{st}\}$ . Then the joint probability density function of  $X(1), \dots, X(N)$  may be written

$$(5.9) \quad f_{X(1), \dots, X(N)}(x_1, \dots, x_N) = \{(2\pi)^N |K_N|\}^{-1} \exp\{-\frac{1}{2}(x, x)_{K_N}\}$$

where  $|K_N|$  is the determinant of  $K_N$ , and the inner product  $(x, x)_{K_N}$  is defined by the right hand side of (5.1). On the other hand, if  $X(1), \dots, X(N)$  satisfy the difference equation  $L_t X(t) = \eta(t)$ , where  $\eta(1), \dots, \eta(N)$  are independent normal random variables with means 0 and variance 1, then

$$(5.10) \quad \begin{aligned} f_{X(1), \dots, X(m), \eta(m+1), \dots, \eta(N)}(x_1, \dots, x_m, y_{m+1}, \dots, y_N) \\ = \{(2\pi)^N |K_m|\}^{-1} \exp[-\frac{1}{2}\{(x, x)_{K_m} + \sum_{j=m+1}^N y_j^2\}]. \end{aligned}$$

Transforming from

$$(X(1), \dots, X(m), \eta(m+1), \dots, \eta(N)) \text{ to } (X(1), \dots, X(N))$$

by the linear transformation  $L_t X(t) = \eta(t)$ ,  $t = m+1, \dots, N$ , it follows from (5.10) that

$$(5.11) \quad \begin{aligned} f_{X(1), \dots, X(N)}(x_1, \dots, x_N) &= \{(2\pi)^N |K_m|\}^{-1} a_0^{N-m} \\ &\cdot \exp\left[-\frac{1}{2}\left\{(x, x)_{K_m} + \sum_{j=m+1}^N |L_t x_t|^2\right\}\right]. \end{aligned}$$

Comparing (5.9) and (5.11) it follows that for any  $N$ -vector  $x$

$$(5.12) \quad (x, x)_{K_N} = (x, x)_{K_m} + \sum_{j=m+1}^N |L_t x_t|^2$$

which is equivalent to (5.4).



To prove (5.6), define the function  $e_j(t)$  by  $e_j(t) = 1$  or 0 according as  $t = j$  or  $t \neq j$ . Since the time series  $X(t)$  is stationary,

$$(5.13) \quad (e_j, e_k)_K = (e_{N-j+1}, e_{N-k+1})_K.$$

For  $1 \leq j, k \leq m$ , defining  $a_j = 0$  for  $j < 0$  or  $j > m$ ,

$$(5.14) \quad \begin{aligned} (e_{N-j+1}, e_{N-k+1})_K &= \sum_{t=m+1}^N L_t(e_{N-j+1})L_t(e_{N-k+1}) \\ &= \sum_{t=m+1}^N a_{j+t-N-1}a_{k+t-N-1} \\ &= \sum_{u=1}^{N-m} a_{j-u}a_{k-u} \end{aligned}$$

while

$$(5.15) \quad \begin{aligned} (e_j, e_k)_K &= d_{jk} + \sum_{t=m+1}^N L_t(e_j)L_t(e_k) \\ &= d_{jk} + \sum_{t=m+1}^N a_{t-j}a_{t-k} \\ &= d_{jk} + \sum_{u=1}^{N-m} a_{u+m-j}a_{u+m-k}. \end{aligned}$$

From (5.13), (5.14), and (5.15), we obtain (5.6).

From (5.4) and (5.6) one may obtain the inverse matrix of the covariance matrix of an autoregressive scheme (see Siddiqui [49] and references cited there).

EXAMPLE 5C: *Autoregressive schemes (continuous parameter)*. We next consider the reproducing kernel Hilbert space corresponding to the covariance kernel of an autoregressive scheme  $X(t)$  observed over a finite interval  $a \leq t \leq b$ .

A continuous parameter stationary time series  $X(t)$  is said to be an autoregressive scheme of order  $m$  if its covariance function  $R(u) = E[X(t)X(t+u)]$  may be written (see Doob [D1], p. 542)

$$(5.16) \quad R(s-t) = \int_{-\infty}^{\infty} \frac{e^{i(s-t)\omega}}{2\pi \left| \sum_{k=0}^m a_k(i\omega)^{m-k} \right|^2} d\omega$$

where the polynomial  $\sum_{k=0}^m a_k z^{m-k}$  has no zeros in the right half of the complex  $z$ -plane. It may be shown that given observations of such a time series over a finite interval  $a \leq t \leq b$ , the corresponding reproducing kernel Hilbert space contains all functions  $h(t)$  on  $a \leq t \leq b$  which are continuously differentiable of order  $m$ . The reproducing kernel inner product is given by

$$(5.17) \quad (h, g)_K = \int_a^b (L_t h)(L_t g) dt + \sum_{j,k=0}^{m-1} d_{j,k} h^{(j)}(a) g^{(k)}(a)$$

where

$$(5.18) \quad L_t h = \sum_{k=0}^m a_k h^{(m-k)}(t)$$

$$(5.19) \quad \{d_{j,k}\}^{-1} = \left\{ \frac{\partial^{j+k}}{\partial v^j \partial u^k} R(t-u) \right\}_{t=u, u=0}.$$

The first and second autoregressive schemes are of particular importance.

A stationary time series  $X(t)$  is said to satisfy a first order autoregressive scheme if it is the solution of a first order linear differential equation whose input is white noise  $\eta'(t)$  (the symbolic derivative of a process  $\eta(t)$  with independent stationary increments):

$$(5.20) \quad (dX/dt) + \beta X = \eta'(t).$$

It should be remarked that from a mathematical point of view (5.20) should be written

$$(5.21) \quad dX(t) + \beta X(t)dt = d\eta(t).$$

Even then, by saying that  $X(t)$  satisfies (5.20) or (5.21) we mean that

$$(5.22) \quad X(t) = \int_{-\infty}^t H(t-s) d\eta(s)$$

where  $H(t-s) = e^{-\beta(t-s)}$  is the one-sided Green's function of the differential operator  $L_t f = f'(t) + \beta f(t)$ .

The covariance function of the stationary time series  $X(t)$  is

$$(5.23) \quad R(t-u) = (1/2\beta)e^{-\beta|u-t|}.$$

The corresponding reproducing kernel Hilbert space  $H(K)$  contains all differentiable functions. The inner product is given by

$$(5.24) \quad (h, g)_K = \int_a^b (h' + \beta h)(g' + \beta g) dt + 2\beta h(a)g(a).$$

More generally, corresponding to the covariance function

$$(5.25) \quad K(s, t) = Ce^{-\beta|s-t|}$$

the reproducing kernel inner product is

$$(5.26) \quad \begin{aligned} (h, g)_K &= \frac{1}{2\beta C} \left\{ \int_a^b (h' + \beta h)(g' + \beta g) dt + 2\beta h(a)g(a) \right\} \\ &= \frac{1}{2\beta C} \int_a^b (h'g' + \beta^2 hg) dt + \frac{1}{2C} \{h(a)g(a) + h(b)g(b)\}. \end{aligned}$$

The random variable  $(h, X)_K$  in  $L_2(X(t), a \leq t \leq b)$  corresponding to  $h(\cdot)$  in

$H(K)$  may be written

$$(5.27) \quad (h, X)_K = \frac{1}{2\beta C} \left\{ \beta^2 \int_a^b h(t)X(t) dt + \int_a^b h'(t) dX(t) \right\} + \frac{1}{2C} \{h(a)X(a) + h(b)X(b)\}.$$

Note that  $X'(t)$  does not exist in any rigorous sense; consequently we write  $dX(t)$  where  $X'(t) dt$  seems to be called for. It can be shown that (5.27) makes sense. In the case that  $h(\cdot)$  is twice differentiable, one may integrate by parts and write

$$(5.28) \quad \int_a^b h'(t) dX(t) = h'(b)X(b) - h'(a)X(a) - \int_a^b X(t)h''(t) dt.$$

A stationary time series  $X(t)$  is said to satisfy a second order autoregressive scheme if it is the solution of a second order linear differential equation whose input is white noise  $\eta'(t)$ :

$$(5.29) \quad (d^2X/dt^2) + 2\alpha(dX/dt) + \gamma^2X = \eta'(t).$$

If  $\omega^2 = \gamma^2 - \alpha^2 > 0$ , the covariance function of the time series is

$$(5.30) \quad R(t-u) = \frac{e^{-\alpha|u-t|}}{4\alpha\gamma^2} \left\{ \cos \omega(u-t) + \frac{\alpha}{\omega} \sin \omega|u-t| \right\}.$$

The corresponding reproducing kernel Hilbert space contains all twice differentiable functions on the interval  $a \leq t \leq b$  with inner product

$$(5.31) \quad (h, g)_K = \int_a^b (h'' + 2\alpha h' + \gamma^2 h)(g'' + 2\alpha g' + \gamma^2 g) dt + 4\alpha\gamma^2 h(a)g(a) + 4\alpha h'(a)g'(a).$$

To write an expression for  $(h, X)_K$ , one uses the same considerations as in (5.27).

Other examples of reproducing kernel Hilbert spaces are given in [39] and [40].

**6. Regression analysis of time series with known covariance function.** The theory of regression analysis (and of the general linear hypothesis) plays a central role in statistical theory. In this section we show how to solve certain standard problems of regression analysis in cases in which the observations possess properties of dependence or continuity. For a discussion of the history and literature of regression analysis the reader is referred to Wold [58].

The classical problem of regression analysis may be posed as follows. Given (i) observations  $X(t)$ ,  $t = 1, \dots, N$ , with known covariance kernel

$$(6.1) \quad K(s, t) = \text{Cov}[X(s), X(t)]$$

and mean value function  $m(t) = E[X(t)]$  of the form

$$(6.2) \quad m(t) = \beta_1 w_1(t) + \dots + \beta_q w_q(t)$$

where  $w_1(\cdot), \dots, w_q(\cdot)$  are known functions, and  $\beta_1, \dots, \beta_q$  are unknown real

numbers, and (ii) a linear function

$$(6.3) \quad \psi(\beta) = \psi_1\beta_1 + \cdots + \psi_q\beta_q$$

of the parameters, where  $\psi_1, \dots, \psi_q$  are known constants. Estimate  $\psi(\cdot)$  by an estimate which (i) is *linear* in the observations in the sense that it is of the form  $\sum_{t=1}^N c_t X(t)$  for some real numbers  $c_1, \dots, c_N$ , (ii) is an *unbiased* estimate of  $\psi(\cdot)$  in the sense, that for all  $\beta = (\beta_1, \dots, \beta_q)$ ,

$$(6.4) \quad E_\beta \left[ \sum_{t=1}^N c_t X(t) \right] = \sum_{t=1}^N c_t m(t) = \sum_{j=1}^q \beta_j \sum_{t=1}^N c_t w_j(t) = \psi(\beta),$$

and (iii) has variance

$$(6.5) \quad \text{Var} \left[ \sum_{t=1}^N c_t X(t) \right] = \sum_{s,t=1}^N c_s K(s,t) c_t$$

equal to the minimum variance of any unbiased linear estimate.

The problem of finding the minimum variance unbiased linear estimate of a linear parametric function  $\psi(\beta)$  can be posed as a problem involving the minimization of a *quadratic form* subject to linear restraints. Define  $K = \{K(s, t)\}$ ,

$$(6.6) \quad W = \begin{bmatrix} w_1(1) & \cdots & w_q(1) \\ \vdots & & \vdots \\ w_1(N) & \cdots & w_q(N) \end{bmatrix}, \quad \psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_q \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix}$$

and let  $c'$  denote the transpose of a (column) vector  $c$ . The unbiasedness condition (6.4) can be stated in matrix form as

$$(6.7) \quad c'W = \psi'.$$

The problem of finding the minimum variance unbiased linear estimate can now be posed as follows: find the vector  $c$  which minimizes the quadratic form  $c'Kc$ , subject to the constraints  $c'W = \psi'$  (compare Bush and Olkin [7]).

**THEOREM 6A.** Let  $K$  be a positive definite  $n \times n$  symmetric matrix,  $W$  be an  $n \times q$  matrix, and  $\psi$  a  $q$ -vector. Assume that

$$(6.8) \quad V = W'K^{-1}W$$

is non-singular. The  $n$ -vector  $c^*$  which minimizes the quadratic form  $c'Kc$  among all  $n$ -vectors  $c$  satisfying  $W'c = \psi$  is given by

$$(6.9) \quad c^* = K^{-1}WV^{-1}\psi$$

and the minimum value of the quadratic form is given by

$$(6.10) \quad c^{*'}Kc^* = \psi'V^{-1}\psi.$$

**PROOF.** One easily verifies that the vector  $c^*$  defined by (6.9) satisfies the restraint  $W'c = \psi$ , and that (6.10) holds. To complete the proof we show that for any  $n$ -vector  $c$  such that  $c'W = \psi'$  it holds that

$$(6.11) \quad c'Kc \geq \psi'V^{-1}\psi.$$

Now for any  $q$ -vector  $z$ , letting  $y = Wz$ ,

$$(6.12) \quad c'Kc \geq \frac{(c'y)^2}{y'K^{-1}y} = \frac{(c'Wz)^2}{z'Vz} = \frac{(\psi'z)^2}{z'Vz}.$$

Taking the supremum of the right side of (6.12) over all  $q$ -vectors  $z$ , one obtains (6.11), since

$$(6.13) \quad \sup_z [(\psi'z)^2/z'Vz] = \psi'V^{-1}\psi.$$

From Theorem 6A, one immediately obtains Theorem 6B.

**THEOREM 6B.** *The minimum variance linear unbiased estimate of a parametric function  $\psi(\beta)$  is*

$$(6.14) \quad \psi^* = c^*X = \psi'V^{-1}(W'K^{-1}X).$$

*The variance of  $\psi^*$  is given by*

$$(6.15) \quad \text{Var} [\psi^*] = c^*Kc^* = \psi'V^{-1}\psi.$$

*In particular, the vector  $\beta^{*'} = (\beta_1^*, \dots, \beta_q^*)$  of minimum variance unbiased linear estimates of  $\beta_1, \dots, \beta_q$  may be written*

$$(6.16) \quad \beta^* = V^{-1}(W'K^{-1}X)$$

*with covariance matrix*

$$(6.17) \quad \{\text{Cov} [\beta_i^*, \beta_j^*]\} = V^{-1}.$$

The foregoing treatment of the problem of regression analysis with known covariance function depended very much on the assumptions that there were only a finite number of observations, and that the matrices  $K$  and  $V$  were non-singular. We now show how to relax these assumptions by using the reproducing kernel Hilbert space representation of a time series. The results we now state include as special cases the results which were first obtained by Grenander ([12], [14]).

Let  $\{X(t), t \in T\}$  be a time series whose covariance kernel  $K(s, t) = \text{Cov} [X(s), X(t)]$  is known and whose mean value function  $m(t) = E[X(t)]$  is only assumed to belong to a known class  $M$ . Let  $H(K)$  be the reproducing kernel Hilbert space corresponding to  $K$ . Assume that  $M$  is a subset of  $H(K)$ . It may be shown that between  $L_2(X(t), t \in T)$  and  $H(K)$  there exists a one-one linear mapping with the following properties: if  $(h, X)_K$  denotes the random variable in  $L_2(X(t), t \in T)$  which corresponds under the mapping to the function  $h$  in  $H(K)$ , then for every  $t$  in  $T$ , and  $h$  and  $g$  in  $H(K)$ ,

$$(6.18) \quad (K(\cdot, t), X)_K = X(t),$$

$$(6.19) \quad E_m[(h, X)_K] = (h, m)_K, \quad \text{for all } m \text{ in } M,$$

$$(6.20) \quad \text{Cov} [(h, X)_K, (g, X)_K] = (h, g)_K.$$

The subscript  $m$  on an expectation operator is written to indicate that the expectation is computed under the assumption that  $m(\cdot)$  is the true mean value function.



If  $T$  is finite, and  $K$  is non-singular, then  $(h, X)_K = h'K^{-1}X$ . For other examples of  $(h, X)_K$ , see Section 5.

We are interested in estimating various functionals  $\psi(m)$  of the true mean value function  $m(\cdot)$  by estimates which (i) are linear in the observations  $\{X(t), t \in T\}$  in the sense that they belong to  $L_2(X(t), t \in T)$ , (ii) are unbiased and (iii) have minimum variance among all linear unbiased estimates. A functional  $\psi(m)$  is said to be linearly estimable if it possesses an unbiased linear estimate  $(g, X)_K$ . Since

$$(6.21) \quad E_m[(g, X)_K] = (g, m)_K = \psi(m), \quad \text{for all } m \text{ in } M$$

it follows that  $\psi(m)$  is linearly estimable if and only if there exists a function  $g$  in  $H(K)$  satisfying (6.21). Now the variance of a linear estimate is given by

$$(6.22) \quad \text{Var}[(g, X)_K] = (g, g)_K.$$

Consequently finding the minimum variance unbiased linear estimate  $\psi^* = (g^*, X)_K$  of  $\psi(m)$  is equivalent to finding that function  $g^*$  in  $H(K)$  which has minimum norm among all functions  $g$  satisfying the restraint (6.21). To find the vector  $g^*$  with minimum norm it suffices to find any vector  $g$  satisfying (6.21). Then the projection

$$(6.23) \quad g^* = E^*[g \mid \bar{M}],$$

of  $g$  onto the smallest Hilbert subspace  $\bar{M}$  containing  $M$ , satisfies (6.21) and has minimum norm among all vectors satisfying (6.21).

**THEOREM 6C.** *The uniformly minimum variance unbiased linear estimate  $\psi^*$  of a linearly estimable function  $\psi(m)$  is given by*

$$(6.24) \quad \psi^* = (E^*[g \mid \bar{M}], X)_K$$

with variance

$$(6.25) \quad \text{Var}[\psi^*] = \|E^*[g \mid \bar{M}]\|_K^2,$$

where  $g$  is any function satisfying (6.21),  $\bar{M}$  is the smallest Hilbert subspace of  $H(K)$  containing  $M$ , and  $E^*[g \mid \bar{M}]$  denotes the projection onto  $\bar{M}$  of  $g$ . In particular, the uniformly minimum variance unbiased linear estimate  $m^*(t)$  of the value  $m(t)$  at a particular point  $t$  of the mean value function  $m(\cdot)$  is given by

$$(6.26) \quad m^*(t) = (E^*[K(\cdot, t) \mid \bar{M}], X)_K$$

since

$$(6.27) \quad m(t) = (K(\cdot, t), m)_K.$$

In the special case that  $M$  consists of all functions  $m(t)$  of the form of (6.2), and the matrix

$$(6.28) \quad V = \begin{bmatrix} (w_1, w_1)_K & \cdots & (w_1, w_q)_K \\ \vdots & & \vdots \\ (w_q, w_1)_K & \cdots & (w_q, w_q)_K \end{bmatrix}$$

is non-singular, then

$$(6.29) \quad \beta^* = V^{-1} \begin{bmatrix} (w_1, X)_K \\ \vdots \\ (w_q, X)_K \end{bmatrix}.$$

One may write an explicit formula for the minimum variance unbiased linear estimate  $\psi^*$  of a linear parametric function  $\psi(\beta) = \psi_1\beta_1 + \cdots + \psi_q\beta_q$  as follows, where  $V_{ij} = (w_i, w_j)_K$ ;

$$(6.30) \quad \psi^* = - \begin{bmatrix} V_{11} \cdots V_{1q} & (X, w_1)_K \\ \vdots & \vdots \\ V_{q1} \cdots V_{qq} & (X, w_q)_K \\ \psi_1 \cdots \psi_q & 0 \end{bmatrix} \div \begin{bmatrix} V_{11} \cdots V_{1q} \\ \vdots \\ V_{q1} \cdots V_{qq} \end{bmatrix}.$$

It should be noted that the proof of Theorem 6C is exactly the same in spirit as the proof of the Gauss-Markov theorem given in Scheffé ([46], p. 14). The point of Theorem 6C is that it enables one to develop a theory of regression analysis and analysis of variance for cases in which one has an infinite number of observations. In particular, we state the analogues of certain basic results on simultaneous confidence intervals (Scheffé [46], p. 68) and hypothesis testing (Scheffé [46], p. 31).

*Hypothesis testing and simultaneous confidence bands for mean value functions.*

If the time series  $X(t)$  is assumed to be normal, or if all linear functionals  $(h, X)_K$  may be assumed to be approximately normally distributed, then one may state a confidence band for the entire mean value function  $m(\cdot)$  as follows. Given a confidence level  $\alpha$ , let  $C_q(\alpha)$  denote the  $\alpha$  percentile of the  $\chi^2$  distribution with  $q$  degrees of freedom; in symbols,  $P[\chi_q^2 \geq C_q(\alpha)] = \alpha$ .

We now show that if the smallest space  $\tilde{M}$  containing all mean value functions has finite dimension  $q$ , then

$$(6.31) \quad m^*(t) - [C_q(\alpha)]^{\frac{1}{2}} \sigma[m^*(t)] \leq m(t) \leq m^*(t) + [C_q(\alpha)]^{\frac{1}{2}} \sigma[m^*(t)]$$

for all  $t$  in  $-\infty < t < \infty$

is a simultaneous confidence band for all values of the mean value function with a level of significance not less than  $\alpha$ ; that is, if  $m(\cdot)$  is the true mean value function then (6.31) holds with a probability greater than or equal to  $\alpha$ .

To prove (6.31) we prove more generally the following theorem.

**THEOREM 6C.** (Simultaneous confidence interval of significance level  $\alpha$  for all estimable functions  $(m, g)$ .) *If  $\tilde{M}$  has dimension  $q$  then for all  $m$  in  $\tilde{M}$*

$$(6.32) \quad P_m \left[ \sup_{g \in H(K)} \frac{|(X, E^*[g | \tilde{M}])_K - (m, g)_K|^2}{\text{Var} [(X, E^*[g | \tilde{M}])_K]} \leq C_q(\alpha) \right] = \alpha.$$

**PROOF.** Let  $w_1, \dots, w_q$  be orthonormal functions which span  $\tilde{M}$ . Then we may write  $m = \beta_1 w_1 + \cdots + \beta_q w_q$  where  $\beta_j = (m, w_j)$  is a function of  $m$ . Further,  $(m, g)_K = \alpha_1 \beta_1 + \cdots + \alpha_q \beta_q$ ,  $(X, E^*[g | \tilde{M}])_K = \alpha_1 \beta_1^* + \cdots + \alpha_q \beta_q^*$ ,

$\text{Var} [(X, E^*[g | \bar{M}])_{\kappa}] = \sum_{j=1}^q \alpha_j^2$ , where  $\alpha_j = (w_j, g)_{\kappa}$  and  $\beta_j^* = (X, w_j)_{\kappa}$ . Next the random variable appearing in (6.32) is equal to

$$(6.33) \quad \sup_{-\infty < \alpha_1, \dots, \alpha_q < \infty} \frac{\left| \sum_{j=1}^q \alpha_j (\beta_j^* - \beta_j) \right|^2}{\sum_{j=1}^q \alpha_j^2} = \sum_{j=1}^q (\beta_j^* - \beta_j)^2$$

which is distributed as  $\chi_q^2$  (compare Scheffé [46], p. 416).

Similarly one may prove the following theorem.

**THEOREM 6D.** *Given a  $q$ -dimensional subspace  $M$  of  $H(K)$ , and a  $q'$ -dimensional subspace  $M'$  of  $M$ , to test the composite null hypothesis  $H_0: m(\cdot) \in M'$ , against the composite alternative hypothesis,  $H_1: m(\cdot) \in M$ , one may use the statistic*

$$(6.34) \quad \Delta = \|m_M^*(t) - m_{M'}^*(t)\|_{\kappa}^2$$

where  $m_M^*(t)$  [ $m_{M'}^*(t)$ ] denotes the minimum variance unbiased linear estimate of  $m(t)$  under the hypothesis  $H_1$  [ $H_0$ ]. Under  $H_0$ ,  $\Delta$  is distributed as  $\chi^2$  with  $q - q'$  degrees of freedom.

In the special case that  $M$  consists of all functions  $m(t)$  of the form (6.2), and  $M'$  consists of all functions in  $M$  for which  $\beta_j = 0$  for  $j = q' + 1, \dots, q$ , then the statistic  $\Delta$  may be written

$$(6.35) \quad \Delta = \sum_{j=q'+1}^q \delta_j$$

where, defining  $V_{ij} = (w_i, w_j)_{\kappa}$ ,

$$(6.36) \quad \delta_j = \frac{|(w_j - E^*[w_j | w_1, \dots, w_{j-1}], X)_{\kappa}|^2}{\|w_j - E^*[w_j | w_1, \dots, w_{j-1}]\|_{\kappa}^2} \\ = \frac{\begin{vmatrix} V_{11} & \dots & V_{1,j-1} & (w_1, X)_{\kappa} \\ \vdots & & \vdots & \\ V_{j1} & \dots & V_{j,j-1} & (w_j, X)_{\kappa} \end{vmatrix}^2}{\begin{vmatrix} V_{11} & \dots & V_{1j} \\ \vdots & & \vdots \\ V_{j1} & \dots & V_{jj} \end{vmatrix} \begin{vmatrix} V_{11} & \dots & V_{1,j-1} \\ \vdots & & \vdots \\ V_{j-1,1} & \dots & V_{j-1,j-1} \end{vmatrix}}$$

The reader may find it illuminating to write out (6.36) in the case that  $q = 2$  and  $q' = 1$ .

*Regression analysis when the covariance function of the observations is only known up to a constant factor.* Suppose that the covariance function of the time series  $\{X(t), t \in T\}$  is of the form

$$\text{Cov} [X(s), X(t)] = \sigma^2 K(s, t)$$

where the kernel  $K(s, t)$  is known and  $\sigma^2$  is an unknown positive constant, and that the mean value function  $m(t) = E[X(t)]$  is known to belong to a set  $M$  which is a subspace (of dimension  $q$ ) of  $H(K)$ , the reproducing kernel Hilbert space corresponding to  $K$ . Theorem 6C continues to hold, except that (6.25) should be replaced by

$$(6.25') \quad \text{Var}_s [\psi^*] = \sigma^2 \|E^*[g | M]\|_{\kappa}^2.$$

The variance of the estimate  $\psi^*$  depends on the unknown parameter  $\sigma^2$ . Therefore one needs to estimate  $\sigma^2$  in order to know  $\text{Var}_\sigma[\psi^*]$ . To discuss the estimation of  $\sigma^2$ , we need to distinguish between the case in which the index set  $T$  is finite and the case in which  $T$  is infinite.

If  $T$  is finite, the time series  $\{X(t), t \in T\}$ , regarded as a function of  $t$ , may be shown to belong to  $H(K)$ . Further if  $n$  is the dimension of  $H(K)$ , then for all possible mean value functions  $m(t)$  and values of  $\sigma^2$

$$\begin{aligned} E[\|X(t) - m(t)\|_K^2] &= n\sigma^2 \\ (6.37) \quad E[\|m^*(t) - m(t)\|_K^2] &= q\sigma^2 \\ E[\|X(t) - m^*(t)\|_K^2] &= (n - q)\sigma^2. \end{aligned}$$

Therefore

$$(6.38) \quad \sigma^{*2} = (n - q)^{-1} \|X(t) - m^*(t)\|_K^2$$

is an unbiased estimate of  $\sigma^2$  (which in the case of normally distributed observations is independent of  $m^*(t)$ ).

If  $T$  is infinite, it is possible to estimate  $\sigma^2$  exactly by forming a sequence of estimates of the form of (6.38) based on a monotone sequence of finite subsets  $\{T_n\}$  of  $T$  whose limit is dense in  $T$ .

**7. The probability density functional of a normal process.** The prediction and regression problems considered in the foregoing have all involved linear estimates chosen according to a criterion expressed in terms of mean square error. Nevertheless the mathematical tools developed continue to play an important role if one desires to employ other criteria of statistical inference. All modern theories of statistical inference take as their starting point the idea of the probability density function of the observations. Thus in order to apply any principle of statistical inference to problems of time series analysis, it is first necessary to develop the notion of the probability density function (or functional) of a stochastic process. In this section we state a result showing how one may write a formula for the probability density function of a stochastic process which is normal.

Given a normal time series  $\{X(t), t \in T\}$  with known covariance function

$$(7.1) \quad K(s, t) = \text{Cov}[X(s), X(t)]$$

and mean value function  $m(t) = E[X(t)]$ , let  $P_m$  be the probability measure induced on the space of sample functions of the time series. Next, let  $m_1$  and  $m_2$  be two functions, and let  $P_1$  and  $P_2$  be the probability measures induced by normal time series with the same covariance kernel  $K$ , and mean value functions equal to  $m_1$  and  $m_2$  respectively. By the Lebesgue decomposition theorem it follows that there is a set  $N$  of  $P_1$ -measure 0 and a non-negative  $P_1$ -integrable function, denoted by  $dP_2/dP_1$ , such that for every measurable set  $B$  of sample functions

$$(7.2) \quad P_2(B) = \int_B (dP_2/dP_1) dP_1 + P_2(BN).$$

If  $P_2(N) = 0$ , then  $P_2$  is absolutely continuous with respect to  $P_1$ , and  $dP_2/dP_1$  is called the probability density function of  $P_2$  with respect to  $P_1$ . Two measures which are absolutely continuous with respect to one another are called *equivalent*. Two measures  $P_1$  and  $P_2$  are said to be *orthogonal* if there is a set  $N$  such that  $P_1(N) = 0$  and  $P_2(N) = 1$ .

It has been proved, independently by various authors under various hypotheses (for references, see [40], Section 4), that two normal probability measures are either equivalent or orthogonal. From the point of view of obtaining an explicit formula for the probability density function, the following formulation of this theorem is useful.

**THEOREM 7A** (Parzen [37], [40]). *Let  $P_m$  be the probability measure induced on the space of sample functions of a time series  $\{X(t), t \in T\}$  with covariance kernel  $K$  and mean value function  $m$ . Assume that either (i)  $T$  is countable or (ii)  $T$  is a separable metric space,  $K$  is continuous, and the stochastic process  $\{X(t), t \in T\}$  is separable. Let  $P_0$  be the probability measure corresponding to the normal process with covariance kernel  $K$  and mean value function  $m(t) = 0$ . Then  $P_m$  and  $P_0$  are equivalent or orthogonal, depending on whether  $m$  does or does not belong to the reproducing kernel Hilbert space  $H(K)$ . If  $m$  belongs to  $H(K)$ , then the probability density functional of  $P_m$  with respect to  $P_0$  is given by*

$$(7.3) \quad f(X, m) = dP_m/dP_0 = \exp \{ (X, m)_K - (\frac{1}{2})(m, m)_K \}.$$

Using the concrete formula for the probability density functional of a normal process provided by (7.3), there is no difficulty in applying the concepts of classical statistical methodology to problems of inference on normal time series. In particular the following theorem may be proved.

**THEOREM 7B.** *Let  $\{X(t), t \in T\}$  be a normal time series, satisfying the assumptions of Theorem 7A with known covariance kernel  $K(s, t) = \text{Cov}[X(s), X(t)]$ , whose mean value function is only assumed to belong to a known class  $M$ . If  $M$  is a finite dimensional subspace of the reproducing kernel space  $H(K)$ , then the maximum likelihood estimate  $m^*(\cdot)$ , defined as that estimate in the space  $M$  of admissible mean value functions such that*

$$(7.4) \quad f(X, m^*) = \max_{m \in M} f(X, m),$$

*exists and is given at each  $t$  in  $T$  by the right hand side of (6.26).*

If  $M$  is an infinite dimensional space, then a maximum likelihood estimate does not exist. This is not too surprising, since  $M$  is not compact in this case. However, an estimate does exist which is the uniformly minimum variance unbiased linear estimate of the value  $m(t)$  at a particular time  $t$  of the mean value function; this estimate is given by (6.26).

The theory of reproducing kernel Hilbert spaces turns out to provide a natural tool for treating problems of minimum variance unbiased estimation (see Parzen [37]). Further work along these lines in the case of normal time series is being done by Ylvisaker ([59]).

**8. Correlation analysis of regression free stationary time series.** In this section, we state some results for discrete parameter time series (a more comprehensive survey is given by Hannan [17]). Many of the results stated may be extended to continuous parameter time series.

We consider a discrete parameter time series  $\{X(t), t = 1, 2, \dots\}$ , with zero means, which is weakly stationary of order 4 in the sense that its covariance function

$$(8.1) \quad R(v) = E[X(t)X(t+v)]$$

and its fourth cumulant function

$$(8.2) \quad \begin{aligned} Q(v_1, v_2, v_3) &= E[X(t)X(t+v_1)X(t+v_2)X(t+v_3)] \\ &\quad - R(v_1)R(v_2-v_3) - R(v_2)R(v_1-v_3) - R(v_3)R(v_1-v_2) \end{aligned}$$

are independent of  $t$ .

**EXAMPLE: Linear Processes.** A discrete parameter time series  $X(t)$  is said to be a linear process, if it may be represented

$$(8.3) \quad X(t) = \sum_{\alpha=-\infty}^{\infty} w(t-\alpha)\eta(\alpha)$$

where  $\sum_{\alpha=-\infty}^{\infty} |w(\alpha)| < \infty$ , and  $\{\eta(\alpha), \alpha = 0, \pm 1, \dots\}$  is a sequence of independent identically distributed random variables with zero means, finite fourth cumulant  $\lambda_4$ , and second cumulant  $\lambda_2$ . A linear process  $X(t)$  is weakly stationary up to order 4, with covariance function, spectral density function, and fourth cumulant function satisfying

$$(8.4) \quad \begin{aligned} R(v) &= \lambda_2 \sum_{\alpha=-\infty}^{\infty} w(\alpha)w(\alpha+v), \quad f(\lambda) = \frac{\lambda_2}{2\pi} \left| \sum_{\alpha=-\infty}^{\infty} w(\alpha)e^{-i\lambda\alpha} \right|^2, \\ Q(v_1, v_2, v_3) &= \lambda_4 \sum_{\alpha=-\infty}^{\infty} w(\alpha)w(\alpha+v_1)w(\alpha+v_2)w(\alpha+v_3), \\ \sum_{u=-\infty}^{\infty} Q(v_1, u, u+v_2) &= \alpha R(v_1)R(v_2), \quad \alpha = \frac{\lambda_4}{(\lambda_2)^2}. \end{aligned}$$

*Correlation analysis* is concerned with estimating the covariance function  $R(v)$ , and the normalized covariance (or correlation) function

$$(8.5) \quad \rho(v) = R(v)/R(0)$$

of a stationary time series.

Given observations  $\{X(t), t = 1, 2, \dots, N\}$ , one can form the sample covariance function, for  $|v| \leq N-1$ ,

$$(8.6) \quad R_N(v) = \frac{1}{N} \sum_{t=1}^{N-|v|} X(t)X(t+|v|)$$

which has mean

$$(8.7) \quad E[R_N(v)] = [1 - (|v|/N)]R(v).$$



As an estimate of  $R(v)$ ,  $R_N(v)$  is biased (although asymptotically unbiased). Consequently if we are interested in estimating  $R(v)$  it may be preferable to take as our estimate

$$R_N^u(v) = [N/(N - |v|)]R_N(v).$$

Many authors have advocated the use of the unbiased estimate  $R_N^u(v)$  in preference to the biased estimate  $R_N(v)$ . However, it appears to me that  $R_N(v)$  is preferable to  $R_N^u(v)$  for two reasons: (i)  $R_N(v)$  is a positive definite function of  $v$ , which is not the case of  $R_N^u(v)$ ; (ii) the mean square error of  $R_N(v)$  as an estimate of  $R(v)$  is in general less than that of  $R_N^u(v)$ . That (i) holds is immediate. That (ii) holds is shown in Parzen [42]. It will be seen that for theoretical purposes it is certainly more useful to consider  $R_N(v)$  rather than  $R_N^u(v)$ .

Using the large of large numbers proved in Parzen ([38], pp. 419-420), one may prove the following theorems on consistency of the sample covariance function.

**THEOREM 8A.** *The sample covariance function of a weakly stationary time series is consistent in quadratic mean, in the sense that, for  $v = 0, 1, \dots$ ,*

$$(8.8) \quad \lim_{N \rightarrow \infty} E |R_N(v) - R(v)|^2 = 0$$

*if the time series is weakly stationary of order 4, and satisfies (for  $v = 0, 1, \dots$ )*

$$(8.9) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{s=0}^{N-1} R^2(s) = 0$$

$$(8.10) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{s=0}^{N-v-1} Q(v, s, v+s) = 0.$$

**THEOREM 8B.** *The sample covariance function of a weakly stationary time series is strongly consistent, in the sense that, for each  $v = 0, 1, \dots$ ,*

$$(8.11) \quad P[\lim_{N \rightarrow \infty} R_N(v) = R(v)] = 1$$

*if the time series is weakly stationary of order 4 and satisfies for positive constants  $C$  and  $q$*

$$(8.12) \quad \frac{1}{N} \sum_{s=0}^{N-1} R^2(s) \leq CN^{-q} \quad \text{for all } N$$

$$(8.13) \quad \frac{1}{N} \left| \sum_{s=0}^{N-v-1} Q(v, s, v+s) \right| \leq CN^{-q} \quad \text{for all } N.$$

In particular, (8.12) and (8.13) hold if it is assumed that

$$(8.14) \quad \sum_{v=-\infty}^{\infty} |R(v)| < \infty$$

$$(8.15) \quad \sum_{v_1, v_2, v_3=-\infty}^{\infty} |Q(v_1, v_2, v_3)| < \infty.$$

We next obtain expressions for the asymptotic covariance of the sample co-

variance function (for proofs of the following theorem see Bartlett [2], [3] or Parzen [36]).

**THEOREM 8C.** *Let  $X(t)$  be a time series weakly stationary of order 4, with absolutely summable covariance and fourth cumulant functions (that is, (8.14) and (8.15) hold). Then the sample covariance function  $R_N(v)$  has asymptotic covariance, for any non-negative integers  $v_1$  and  $v_2$ ,*

$$(8.16) \quad \lim_{N \rightarrow \infty} N \text{Cov} [R_N(v_1), R_N(v_2)] = D(v_1, v_2)$$

where we define

$$(8.17) \quad D(v_1, v_2) = \sum_{u=-\infty}^{\infty} \{R(u)R(u+v_2-v_1) + R(u)R(u+v_2+v_1) + Q(v_1, u, u+v_2)\}.$$

For a linear process with spectral density function  $f(\cdot)$

$$(8.18) \quad D(v_1, v_2) = 4\pi \int_{-\pi}^{\pi} \cos \lambda v_1 \cos \lambda v_2 f^2(\lambda) d\lambda + \alpha \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \cos \lambda v_1 \cos \lambda v_2 f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2.$$

In particular, the variance of  $R_N(v)$  is approximately given by

$$(8.19) \quad \text{Var} [R_N(v)] = \frac{4\pi}{N} \int_{-\pi}^{\pi} \cos^2 \lambda v f^2(\lambda) d\lambda + \frac{\alpha}{N} R^2(v) \geq \frac{2+\alpha}{N} R^2(v).$$

The mean square error of  $R_N(v)$  as an estimate of  $R(v)$  is given by

$$(8.20) \quad E |R_N(v) - R(v)|^2 = \text{Var} [R_N(v)] + \left(\frac{v}{N}\right)^2 R^2(v).$$

It was empirically observed by M. G. Kendall that the sample covariance function (traditionally called the *observed correlogram*) fails to damp down to 0 for increasing values of  $v$ , although the true covariance function  $R(v)$  does damp down to 0 as  $v$  tends to  $\infty$ . This fact is borne out theoretically by (8.19) and (8.20), which show that the coefficient of mean square error  $E |R_N(v) - R(v)|^2 / R^2(v)$  is of the order of  $1/N$  for all lags  $v$  of the sample covariance function.

One may state in a variety of ways conditions under which the sample covariance function  $R_N(v)$  is asymptotically normal in the sense that for every choice of lags  $v_1, \dots, v_k$  and real numbers  $u_1, \dots, u_k$ ,

$$(8.21) \quad E[\exp i\{u_1 N^{1/2}(R_N(v_1) - E[R_N(v_1)]) + \dots + u_k N^{1/2}(R_N(v_k) - E[R_N(v_k)])\}] \rightarrow \exp \left[ -\frac{1}{2} \left\{ \sum_{i,j=1}^k u_i D(v_i, v_j) u_j \right\} \right]$$

as  $N \rightarrow \infty$  (see Walker [54], Lomnicki and Zaremba [29], Parzen [35]). In particular, (8.21) holds if  $X(t)$  is a linear process.

As an estimate of the correlation function  $\rho(v)$  we take the sample correlation function

$$(8.22) \quad \rho_N(v) = R_N(v)/R_N(0).$$

We do not discuss here the question of the best definition of the correlation function for short series. For a discussion of this problem, and references to the literature, see Weinstein [55].

By standard large sample statistical theory one readily obtains, from (8.21) and (8.16), the following theorem.

**THEOREM 8D.** *If  $X(t)$  is a linear process, then  $\rho_N(v)$  is asymptotically normal with asymptotic covariances satisfying, as  $N \rightarrow \infty$ ,*

$$(8.23) \quad NE[\rho_N(v_1) - \rho(v_1), \rho_N(v_2) - \rho(v_2)] \rightarrow d(v_1, v_2)$$

where we define

$$(8.24) \quad d(v_1, v_2) = 4\pi \int_{-\pi}^{\pi} d\lambda \bar{f}^2(\lambda) \{\cos \lambda v_1 - \rho(v_1)\} \{\cos \lambda v_2 - \rho(v_2)\},$$

$$(8.25) \quad \bar{f}(\lambda) = \frac{f(\lambda)}{R(0)}.$$

**REMARK.** It should be noted that while the variance of the sample covariance function  $R_N(v)$  of a linear process depends on  $\alpha$ , the variance of the sample correlation function  $\rho_N(v)$  does not.

**PROOF.** Using only the first few terms of the Taylor series expansion one obtains that

$$(8.26) \quad \frac{x}{y} - \frac{x_0}{y_0} = \frac{y}{y_0^2} \{(y - y_0)x_0 - (x - x_0)y_0\} + O(|x - x_0|^2 + |y - y_0|^2)$$

Consequently, if  $X_n$ ,  $Y_n$ , and  $Z_n$  are sequences of random variables, and  $x_0$ ,  $y_0$ , and  $z_0$  are constants, such that

$$n^{1/2}(X_n - x_0), \quad n^{1/2}(Y_n - y_0), \quad n^{1/2}(Z_n - z_0)$$

are jointly asymptotically normal it follows that

$$n^{1/2} \left( \frac{X_n}{Y_n} - \frac{x_0}{y_0} \right), \quad n^{1/2} \left( \frac{Z_n}{Y_n} - \frac{z_0}{y_0} \right)$$

are jointly asymptotically normal with asymptotic covariance satisfying

$$(8.27) \quad \begin{aligned} y_0^4 n \text{Cov} \left[ \frac{X_n}{Y_n} - \frac{x_0}{y_0}, \frac{Z_n}{Y_n} - \frac{z_0}{y_0} \right] &\rightarrow x_0 z_0 E[(Y_n - y_0)^2] \\ &+ y_0^2 E[(X_n - x_0)(Z_n - z_0)] - x_0 y_0 E[(Y_n - y_0)(Z_n - z_0)] \\ &\quad - z_0 y_0 E[(Y_n - y_0)(X_n - x_0)]. \end{aligned}$$

Applying these results to the present case it follows that the sample correlations

$\rho_N(v)$  are asymptotically normal with asymptotic covariances satisfying

$$\begin{aligned} R^4(0)N \text{Cov} [\rho_N(v_1) - \rho(v_1), \rho_N(v_2) - \rho(v_2)] \\ (8.28) \quad = R(v_1)R(v_2)D(0, 0) + R^2(0)D(v_1, v_2) - R(0)R(v_1)D(0, v_2) \\ \quad \quad - R(0)R(v_2)D(0, v_1). \end{aligned}$$

From (8.28) and (8.18), one obtains (8.24).

We are now in a position to obtain confidence intervals for, or test hypotheses about, a correlation coefficient  $\rho(v)$ . From Theorem 8D it follows that the sample correlation coefficient  $\rho_N(v)$  may be regarded as being normally distributed with mean  $\rho(v)$  and variance equal to  $d(v)/N$  where we define

$$(8.29) \quad d(v) = 4\pi \int_{-\pi}^{\pi} d\lambda \tilde{f}^2(\lambda) \{\cos \lambda v - \rho(v)\}^2.$$

Now  $d(v) \leq 16\pi \int_{-\pi}^{\pi} d\lambda \tilde{f}^2(\lambda)$ ; further, for large values of  $v$ , approximately  $d(v) = 2\pi \int_{-\pi}^{\pi} d\lambda \tilde{f}^2(\lambda)$ . One thus sees that in order to obtain bounds for  $d(v)$  one must have a knowledge of the quantity

$$(8.30) \quad d = 2\pi \int_{-\pi}^{\pi} d\lambda \frac{f^2(\lambda)}{R^2(0)} = \frac{1}{R^2(0)} \sum_{v=-\infty}^{\infty} R^2(v).$$

In the study of both correlation analysis and spectral analysis of stationary time series it will be found that the quantity  $d$  arises frequently as information which one requires about the time series under consideration in order to carry out various statistical procedures. A satisfactory estimate of  $d$  from observations  $\{X(t), t = 1, 2, \dots, N\}$  is provided by

$$(8.31) \quad d_N = \frac{1}{2R_N^2(0)} \sum_{v=-(N-1)}^{N-1} R_N^2(v).$$

If one does not desire to compute the sample covariance function for all  $v = 0, 1, \dots, N$  then one may take, for any  $\theta$  in  $0 < \theta \leq 1$ ,

$$(8.32) \quad d_{N,(\theta N)} = \frac{1}{(1 + 2\theta - \theta^2)R_N^2(0)} \sum_{v=-(\theta N)}^{(\theta N)} R_N^2(v)$$

as an estimate of  $d$ . The properties of the estimates  $d_N$  and  $d_{N,(\theta N)}$  have been extensively investigated by Lomnicki and Zaremba [29]; among other things they show that  $d_N$  is a consistent estimate of  $d$  which in the case of a linear process has an asymptotic variance not dependent on the residuals  $\{\eta(\alpha)\}$ .

An alternate approach to the problem of investigating the mechanism generating a time series is to attempt to fit the time series by a finite parameter scheme (such as an autoregressive scheme or a moving average scheme). Here we consider only the problem of fitting an autoregressive scheme which has the most developed theory (for recent work on fitting moving average schemes, see Durbin [10]).

**THEOREM 8E.** *In order that a stationary time series  $X(t)$  with covariance func-*

tion  $R(v)$ , satisfy the autoregressive scheme of order  $m$ ,

$$(8.33) \quad X(t) = a_1 X(t-1) + \cdots + a_m X(t-m) + \eta(t)$$

where  $\eta(t)$  are a sequence of orthogonal random variables (with common variance  $\sigma^2$ ) representing the innovation at time  $t$  so that

$$(8.34) \quad E[X(s)\eta(t)] = 0 \quad \text{for } s < t$$

it is necessary and sufficient that the covariance function  $R(v)$  satisfy the difference equation

$$(8.35) \quad R(u) = a_1 R(u-1) + \cdots + a_m R(u-m) \quad \text{for } u > 0$$

while for  $u = 0$

$$(8.36) \quad R(0) = a_1 R(1) + \cdots + a_m R(m) + \sigma^2.$$

REMARK. Equations (8.35) are called the Yule-Walker equations, after G. Udny Yule and Sir Gilbert Walker who first obtained relations of this kind (see Wold [58], especially pp. 104-5 and pp. 140-146).

PROOF. Verify that (8.33) and (8.35) are each equivalent to the assertion that the minimum means square error linear predictor of  $X(t)$ , given its infinite past, depends only on the finite past  $X(t-1), \dots, X(t-m)$ ; in symbols, for all  $t$

$$(8.36) \quad \begin{aligned} E^*[X(t) \mid X(t-1), \dots, X(t-m), \dots] \\ = a_1 X(t-1) + \cdots + a_m X(t-m). \end{aligned}$$

We may use the fact that the covariance of a stationary autoregressive scheme satisfies the difference equation (8.35) to obtain expressions for the constants  $a_1, \dots, a_m$  in terms of correlations; (8.35) with  $u = 1, \dots, m$  yields  $m$  equations which may be written in matrix form

$$(8.37) \quad \begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(m-1) \\ \rho(1) & \rho(0) & \cdots & \rho(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(m-1) & \rho(m-2) & \cdots & \rho(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(m) \end{bmatrix}.$$

Consistent asymptotically normal estimates  $a_1^{(N)}, \dots, a_m^{(N)}$ , of  $a_1, \dots, a_m$  respectively, may be obtained from observations  $\{X(t), t = 1, 2, \dots, N\}$  by forming consistent asymptotically normal estimates  $\rho_N(v)$  of  $\rho(v)$  and defining  $a_1^{(N)}$  to be the solutions of

$$\begin{bmatrix} \rho_N(0) & \rho_N(1) & \cdots & \rho_N(m-1) \\ \rho_N(1) & \rho_N(0) & \cdots & \rho_N(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_N(m-1) & \rho_N(m-2) & \cdots & \rho_N(0) \end{bmatrix} \begin{bmatrix} a_1^{(N)} \\ a_2^{(N)} \\ \vdots \\ a_m^{(N)} \end{bmatrix} = \begin{bmatrix} \rho_N(1) \\ \rho_N(2) \\ \vdots \\ \rho_N(m) \end{bmatrix}.$$

It may be shown, using standard techniques of large sample statistical theory,

that if the estimates  $R_N(v)$  satisfy (8.21), then the estimates  $a_i^{(N)}$  satisfy

$$(8.39) \quad E[\exp i\{u_1(N^{\frac{1}{2}})(a_1^{(N)} - a_1) + \dots + u_m N^{\frac{1}{2}}(a_m^{(N)} - a_m)\}] \rightarrow \exp \left[ -\frac{1}{2} \sum_{i,j=1}^m u_i \sigma^2 C_{ij} u_j \right]$$

where  $\{C_{ij}\}$  is the inverse matrix of the  $m$  by  $m$  matrix whose  $(i, j)$ th entry is  $\rho(i-j)$ . As an estimate  $\{C_{ij}^{(N)}\}$  of  $\{C_{ij}\}$  one may take the inverse matrix of  $\{\rho_N(i-j)\}$ , and as an estimate of  $\sigma^2$  one may take

$$(8.40) \quad \sigma_N^2 = \frac{1}{N-m} \sum_{t=m+1}^N \{X(t) - a_1^{(N)}X(t-1) - \dots - a_m^{(N)}X(t-m)\}^2.$$

In words, (8.39) says that the usual theorems of regression analysis apply asymptotically to the problem of estimating the autoregressive coefficients, even though the regression functions  $X(t-1), \dots, X(t-m)$  represent lagged values of the observed time series  $X(t)$ . This fact was first shown by Mann and Wald [31] whose paper is a fundamental contribution to the theory of time series analysis.

To prove (8.39), we write (8.37) and (8.38) in alternate form as follows. Define  $a_0 = a_0^{(N)} = 1$ . Then, for  $i = 1, 2, \dots, m$

$$(8.37') \quad \sum_{j=0}^m a_j \rho(i-j) = 0$$

$$(8.38') \quad \sum_{j=0}^m a_j^{(N)} \rho_N(i-j) = 0.$$

Therefore for  $i = 1, \dots, m$

$$(8.41) \quad \sum_{j=0}^m \rho_N(i-j) \{a_j^{(N)} - a_j\} = \sum_{j=0}^m a_j \{\rho(i-j) - \rho_N(i-j)\}.$$

From (8.41) one may deduce (8.39).

EXAMPLE. Let us write out the foregoing formulas for the case of an autoregressive scheme of order 2. Then (8.38) may be written

$$(8.42) \quad \begin{aligned} a_1^{(N)} + a_2^{(N)} \rho_N(1) &= \rho_N(1) \\ a_1^{(N)} \rho_N(1) + a_2^{(N)} &= \rho_N(2). \end{aligned}$$

The estimates  $a_1^{(N)}$  and  $a_2^{(N)}$  are explicitly given by

$$(8.43) \quad \begin{aligned} a_1^{(N)} &= \frac{\rho_N(1) \{1 - \rho_N(2)\}}{1 - \rho_N^2(1)} \\ a_2^{(N)} &= \frac{\rho_N(2) - \rho_N^2(1)}{1 - \rho_N^2(1)}. \end{aligned}$$

The estimated covariance matrix of  $\{\text{Cov}[a_i^{(N)}, a_j^{(N)}]\}$  is given by



$$(8.44) \quad \{\text{Cov} [a_i^{(N)}, a_j^{(N)}]\}_{\text{est}} = \frac{1}{N} \sigma_N^2 \begin{bmatrix} 1 & \rho_N(1) \\ \rho_N(1) & 1 \end{bmatrix}^{-1}.$$

To test the null hypothesis that the time series obeys an autoregressive scheme or order 1 against the alternative hypothesis that it obeys an autoregressive scheme of order 2 one uses the statistic

$$(8.45) \quad \delta = \frac{|a_2^{(N)}|^2}{\text{Var} [a_2^{(N)}]} = \frac{N\{\rho_N(2) - \rho_N^2(1)\}^2}{\sigma_N^2\{1 - \rho_N^2(1)\}}$$

which under the null hypothesis is distributed as  $\chi^2$  with 1 degree of freedom. One may similarly give a test of the null hypothesis that the time series obeys an autoregressive scheme of order  $q'$  against the alternative hypothesis that it obeys an autoregressive scheme of order  $q$  (greater than  $q'$ ).

For an excellent review of both the small and large sample theory of goodness of fit tests for autoregressive schemes, we refer the reader to the monograph by E. J. Hannan [17]. For references to recent work on explosive stochastic difference equations, see Rao [44a].

#### REFERENCES

- [1] N. ARONSZAJN, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, Vol. 68 (1950), pp. 337-404.
- [2] M. S. BARTLETT, "On the theoretical specification and sampling properties of auto-correlated time series," *J. Roy. Stat. Soc., Suppl.*, Vol. 8 (1946), pp. 27-41.
- [3] M. S. BARTLETT, *An Introduction to Stochastic Processes*, Cambridge, 1955.
- [4] R. B. BLACKMAN AND J. W. TUKEY, *The Measurement of Power Spectra from the Point of View of Communications Engineering*, New York, Dover Publications, 1959.
- [5] H. W. BODE AND C. E. SHANNON, "A simplified derivation of linear least squares smoothing and prediction theory," *Proc. I.R.E.*, Vol. 38 (1950), pp. 417-425.
- [6] C. E. P. BROOKS AND N. CARRUTHERS, *Handbook of Statistical Methods in Meteorology*, London, H. M. Stationary Office, 1953.
- [7] K. A. BUSH AND I. OLKIN, "Extrema of quadratic forms with applications to statistics," *Biometrika*, Vol. 46 (1959), pp. 483-486.
- [8] LOLA S. DEMING, "Selected bibliography of statistical literature, 1930-1957: II. Time series," *J. Res. Nat. Bur. Standards*, Vol. 64B (1960), pp. 69-76.
- [9] J. L. DOOB, *Stochastic Processes*, New York, John Wiley and Sons, 1953.
- [10] J. DURBIN, "Efficient estimation of parameters in moving average models," *Biometrika*, Vol. 46 (1959), pp. 306-316.
- [11] R. A. FISHER, "Tests of significance in harmonic analysis," *Proc. Roy. Soc., Ser. A*, Vol. 125 (1929), p. 54.
- [12] U. GRENANDER, "Stochastic processes and statistical inference," *Arkiv fur Matematik*, Vol. 1 (1950), pp. 195-277.
- [13] U. GRENANDER AND M. ROSENBLATT, *Statistical Analysis of Stationary Time Series*, New York, John Wiley and Sons, 1957.
- [14] U. GRENANDER AND G. SZEGO, *Toeplitz Forms and their Applications*, Berkeley, University of California Press, 1958.
- [15] T. N. E. GREVILLE, "Some applications of the pseudo inverse of a matrix," *SIAM Review*, Vol. 2 (1960), pp. 15-22.
- [16] P. HALMOS, *Introduction to Hilbert Space*, Chelsea, New York, 1951.
- [17] E. J. HANNAN, *Time Series Analysis*, Methuen, London, 1960.

- [18] H. O. HARTLEY, "Tests of significance in harmonic analysis," *Biometrika*, Vol. 36 (1949), pp. 194-201.
- [19] M. P. HEBLE, "Some problems in time series analysis," *J. Math. and Mechanics*, Vol. 9 (1960), pp. 951-987.
- [20] G. M. JENKINS, "General considerations in estimation of spectra," *Technometrics*, Vol. 3 (1961), pp. 133-166.
- [21] K. KARHUNEN, "Über linear methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae*, I. Math.-Physica, No. 37 (1947).
- [22] M. G. KENDALL, "On the analysis of oscillatory time series," *J. Roy. Stat. Soc.*, Vol. 106 (1945), p. 91.
- [23] M. G. KENDALL, *Advanced Theory of Statistics*, Vol. II, London, Charles Griffin, 1946.
- [24] A. KHINTCHINE, "Korrelationstheorie der stationären stochastischen Prozesse," *Math. Ann.*, Vol. 109 (1933), pp. 604-615.
- [25] A. KOLMOGOROV, "Interpolation and extrapolation," *Bull. Acad. Sci. U.R.S.S.*, Ser. Math., Vol. 5 (1941), pp. 3-14.
- [26] J. H. LANNING AND R. H. BATTIN, *Random Processes in Automatic Control*, New York, McGraw Hill Book Co., 1956.
- [27] M. LOÈVE, "Fonctions aléatoires du second ordre," Suplement to P. Lévy, *Processus Stochastiques et Mouvement Brownien*, Paris, Gauthier-Villars, 1948.
- [28] M. LOÈVE, *Probability Theory*, New York, D. Van Nostrand, 1955.
- [29] Z. A. LOMNICKI AND S. K. ZAREMBA, "On some moments and distributions occurring in the theory of linear stochastic processes," *Monatshefte für Mathematik*, Part I, Vol. 61 (1957), pp. 318-358; Part II, Vol. 63 (1959), pp. 128-168.
- [30] M. S. LONGUET-HIGGINS, "On the intervals between successive zeros of a random function," *Proc. Roy. Soc.*, Ser. A, Vol. 246 (1958), pp. 99-118.
- [31] H. B. MANN AND A. WALD, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, Vol. 11 (1943), pp. 173-220.
- [32] D. MIDDLETON, *Statistical Communication Theory*, New York, McGraw Hill Book Co., 1960.
- [33] P. A. P. MORAN, "Random processes in economic theory and analysis," *Sankhyā*, Vol. 21 (1959), pp. 99-126.
- [34] J. NEYMAN, "Current problems of mathematical statistics," *Proceedings of the International Congress of Mathematicians*, Amsterdam, 1954, Volume 1.
- [35] E. PARZEN, "A central limit theorem for multilinear stochastic processes," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 252-255.
- [36] E. PARZEN, "On consistent estimates of the spectrum of a stationary time series," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 329-348.
- [37] E. PARZEN, "Statistical inference on time series by Hilbert space methods, I," Department of Statistics, Stanford University, Technical Report No. 23, January 2, 1959.
- [38] E. PARZEN, *Modern Probability Theory and its Applications*, John Wiley and Sons, 1960.
- [39] E. PARZEN, "A new approach to the synthesis of optimal smoothing and prediction systems," Department of Statistics, Stanford University, Technical Report No. 34, July 15, 1960. (To be published in the *Proceedings of a Symposium on Optimization*, edited by R. Bellman.)
- [40] E. PARZEN, "Regression analysis of continuous parameter time series, II," Department of Statistics, Stanford University, Technical Report No. 35, July 28, 1960. (To be published in the *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability Theory*, edited by J. Neyman.)
- [41] E. PARZEN, "Mathematical considerations in estimation of spectra," *Technometrics*, Vol. 3 (1961), pp. 167-190.
- [42] E. PARZEN, "Estimating the covariance function of a stationary time series," to be published.

- [42a] E. PARZEN AND G. R. HEXT, "A Bibliography of American Publications on Stochastic Processes and Time Series Analysis Published in the years 1900-1959 (Preliminary Report)," Stanford University Technical Report, 1961.
- [43] V. S. PUGACHEV, "A method of solving the basic integral equation of statistical theory of optimum systems in finite form" (in Russian), *J. Applied Math. and Mechanics*, Vol. 23 (1959), pp. 3-14; English translation, pp. 1-16.
- [44] V. S. PUGACHEV, *Theory of Random Functions and its Application to Automatic Control Problems* (in Russian), 2nd ed., State Publishing House of Theoretical Technical Literature (Gostekhizdat), Moscow, 1960.
- [44a] M. M. RAO, "Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations," *Ann. Math. Stat.*, Vol. 32 (1961), pp. 195-218.
- [45] M. ROSENBLATT, "Statistical analysis of stochastic processes with stationary residuals," *Probability and Statistics*, edited by U. Grenander, John Wiley and Sons, New York, 1959, pp. 246-273.
- [46] H. SCHEFFÉ, *Analysis of Variance*, John Wiley and Sons, New York, 1959.
- [47] A. SCHUSTER, "On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena," *Terr. Mag.*, Vol. 3 (1898), p. 13.
- [48] A. SCHUSTER, "On the periodicities of sunspots," *Phil. Trans.*, Ser. A., Vol. 206 (1906), p. 69.
- [49] M. M. SIDDIQUI, "On the inversion of the sample covariance matrix in a stationary autoregressive process," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 585-588.
- [50] D. SLEPIAN, "First passage problem for a particular Gaussian process," *Ann. Math. Stat.*, Vol. 32 (1961), pp. 610-612.
- [51] E. SLUTSKY, "The summation of random causes as the source of cyclic processes," *Econometrika*, Vol. 5 (1937), p. 105 (translation of a 1927 paper).
- [52] C. T. STRIEBEL, "Efficient estimation of regression parameters for certain second order stationary processes," Ph.D. dissertation, University of California, 1960.
- [53] J. W. TUKEY, "An introduction to the measurement of spectra," *Probability and Statistics*, edited by U. Grenander, John Wiley and Sons, New York, 1959, pp. 300-330.
- [54] A. M. WALKER, "The asymptotic distribution of serial correlation coefficients for autoregressive processes with dependent residuals," *Proc. Camb. Phil. Soc.*, Vol. 50 (1954), pp. 60-64.
- [55] ABBOTT S. WEINSTEIN, "Alternative definitions of the serial correlation coefficient in short autoregressive sequences," *J. Amer. Stat. Ass.*, Vol. 53 (1958), pp. 881-892.
- [56] N. WIENER, "Generalized harmonic analysis," *Acta. Math.*, Vol. 55 (1930), p. 117.
- [56a] N. WIENER, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, John Wiley and Sons, New York, 1949.
- [57] H. WOLD, *Demand Analysis*, John Wiley and Sons, New York, 1953.
- [58] H. WOLD, *A Study in the Analysis of Stationary Time Series*, Uppsala, 2nd ed., 1954 (with an appendix by Peter Whittle).
- [59] D. YLVISAKER, "On a class of covariance kernels admitting a power series expansion," to be published.
- [60] G. U. YULE, "On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers," *Phil. Trans.*, Ser. A, Vol. 226 (1927), p. 267.
- [61] L. A. ZADEH AND J. R. RAGAZZINI, "An extension of Wiener's theory of prediction," *J. Applied Physics*, Vol. 21 (1950), pp. 645-655.
- [62] L. A. ZADEH AND J. R. RAGAZZINI, "Optimum filters for the detection of signals in noise," *Proc. I. R. E.*, Vol. 40 (1952), pp. 1223-1231.

## SOME MODEL I PROBLEMS OF SELECTION

BY E. L. LEHMANN<sup>1</sup>

*University of California, Berkeley*

**1. Summary.** There are given a populations  $\Pi_1, \dots, \Pi_n$ , of which we wish to select a subset. The quality of the  $i$ th population is characterized by a real-valued parameter  $\theta_i$ , and a population is said to be

(1) *positive (or good)* if  $\theta_i \geq \theta_0 + \Delta$ ,

(2) *negative (or bad)* if  $\theta_i \leq \theta_0$ ,

where  $\Delta$  is a given positive constant and  $\theta_0$  is either a given number or a parameter that may be estimated. A number of optimum properties of selection procedures are defined (Section 3) and it is shown that for some of these, the optimum procedure selects  $\Pi_i$  when

(3)  $T_i \geq C_i$ ,

where  $T_i$  is a suitable statistic, the distribution of which depends only on  $\theta_i$ , and where  $C$  is a suitable constant. (Sections 4 and 6.) Applications are given to distributions with monotone likelihood ratio in the case that  $\theta_0$  is known (Sections 5 and 6), and to normal distributions when instead observations on  $\theta_0$  are included in the experiment (Sections 10 and 11).

**2. Introduction.** An important class of classification problems is concerned with *selection*, that is with the classification of items into a superior category (the selected items) and an inferior one. We shall not be concerned here with more general classification procedures which would divide the items into possibly more than two categories. Selection problems have been treated in many different formulations. A basic distinction is that corresponding to Models I and II in the analysis of variance. In Model I, the items being classified are considered fixed; only the observations made on each item are random. In Model II, on the other hand, the items themselves are drawn at random from some population and would therefore change under a replication of the experiment. Model II problems have been treated recently, among others by Z. W. Birnbaum [1], Birnbaum and Chapman [2], T. W. Anderson [3], Cochran [4], Finney [5, 6], Davies [7], Curnow [8] and Dunnett [9]. For the related problem of the rejection of outliers, see for example [10]. We shall in the present paper be concerned only with Model I.

We shall assume therefore that a number of varieties, treatments, production

---

Received February 18, 1960, revised July 13, 1961.

<sup>1</sup> This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43 and Nonr-2842(00)). This paper in whole or in part may be reproduced for any purpose of the United States Government.

methods, etc. (in general we shall speak of *populations*) are at our disposal. Their quality is characterized by a measure which we shall assume to be scalar. From the available set we wish to make a selection, selecting as far as possible the best ones.

It is useful to consider two cases according to the size of the group to be selected.

PROBLEM 1. A first possibility is that we wish to select only a single population (if possible the best one): the variety to be planted, the production method we are going to adopt, etc. As a slight generalization we may wish to select a fixed number, say two or three. We may have a fixed number of prizes or fellowships to award, or we may not wish to put all eggs into one basket.

PROBLEM 2. In the second case the group size is variable and is determined by the observations. This arises for example when we wish to select all worthwhile treatments or if we want to be reasonably sure that the selected group contains the best treatment.

PROBLEM 3. There is finally the intermediate possibility that the group size is variable but has a fixed upper limit. It may for example be desirable to investigate all treatments that appear promising but budget restrictions may limit the research program to the investigation of at most three treatments.

The traditional formal treatment of the class of problems described here, which has always been recognized as inadequate, is through tests of homogeneity (as for example in the analysis of variance). The only question answered by such a test is whether there is any difference at all among the available populations.

The first step toward a more realistic formulation is due to Mosteller [11] who gave a procedure for testing the hypothesis of homogeneity against the *slippage alternatives* that exactly one of the populations has slipped to the right and for deciding, in case of rejection of the hypothesis, just which of the populations slipped. Mosteller's paper was at least a partial answer to such an urgent need that, in spite of his warnings regarding certain inadequacies in the formulation, it inspired a large literature on slippage tests. At the same time, it led to further clarification of the issues. The first completely satisfactory proposal for dealing with a problem of type 2 above was made by Paulson [12], while problem 1 was formulated and essentially solved<sup>2</sup> by Bahadur [13].

Most of the literature on selection problems so far has been concerned with the definition of suitable procedures, an evaluation of their performance characteristics and the determination of the sample size. An optimum theory was developed for problem 1 by Bahadur in [13] and by Bahadur and Goodman in [14]. An optimum property of a slippage test, with reference only to slippage alternatives, was first proved by Paulson [15]. His proof was applied to other problems, was generalized and simplified in papers by Doornbos and Prins [16], Kudo [17], Pfanzagl [18], Ramachandran and Khatri [19], Truax [20], and Karlin and Truax [21]. Finally, contributions toward optimum properties of

---

<sup>2</sup> For the nonsequential case which is the only one considered here.

procedures for problem 2 were made by Gupta [22], Robbins [23] and Seal [24], [25], [26]. In the present paper, we shall be concerned with optimum procedures for certain cases of problem 2.

It is useful to introduce here another distinction according to the definition of the quality of a population.

**A.** In the simplest case, the quality is defined in absolute terms. If a number of new treatments are being compared with a standard treatment, it may happen that the latter has been observed so extensively that its effect can be taken as known. A treatment is then "good" if it is better (or sufficiently much better) than the standard. Another example is furnished by the selection of binomial populations, where success probabilities are compared with the "pure chance" value  $p = \frac{1}{2}$ , a probability  $p_i$  being considered as good if it exceeds this value or exceeds it by at least a given amount.

**B.** Usually, in the comparison of new treatments with a standard, it is of course better not to treat the standard as known but to let it participate in the experiment as a control. A new treatment is then "good" if it compares favorably with the control, the effect of which is also determined by the experiment.

**C.** Comparisons are not always relative to a standard or control. If a new product is being developed, it may be a question of selecting the most promising of a number of variants or a number of production methods. In such a case, each population must be compared with the totality of the remaining populations. A population may then be considered as "good" if it is (sufficiently much) better than the average of the remaining populations or if it does not fall too much below the best one. In the present paper, only problem A and B will be considered.

We mention in conclusion that the applications of selection theory are even wider than may appear at first: The emphasis instead of on selection may be on elimination. Thus we may wish to eliminate those regression coefficients or interactions, which can safely be neglected or those observations that represent gross errors. In the latter context slippage procedures, that is, procedures derived under the assumption of at most one "outlier" were proposed and their disadvantages discussed quite early by Pearson and Chandrasekhar [27].

**3. Formulation of the problem.** As in the Neyman-Pearson theory of hypothesis testing, there are two possible sources of error in any set of selections. There is the possibility of *false positives*, that is, populations which are selected although they are negative ( $\leq \theta_0$ ), and of *false negatives*, that is, populations which are not selected although they are positive ( $\geq \theta_0 + \Delta$ ). Instead of on false negatives we shall focus attention on *true positives*, that is, on those positive populations which are included in the selected group. This is analogous to the replacement of the consideration of an error of the second kind by that of power in the Neyman-Pearson theory.

Roughly speaking, it is the aim of a selection procedure to seek out the true positives while holding false positives to a minimum.



For measuring how well a procedure carries out its task of identifying the positive populations, a number of criteria are available.

(a) The expected number of true positives.

(b) The expected proportion of true positives, that is, the quantity (a) divided by the total number of positives.

These criteria are appropriate if it is desired to include in the selected group as many of the positive populations as possible.

(c) The probability of at least one true positive.

(d) The probability of including in the selected group the best population (that is, the population with the largest  $\theta$ -value), provided it is positive.

These two criteria may be appropriate if the selection is only a step in a scheme, of which the eventual aim is the selection of a single population.<sup>3</sup>

(e) The probability of including all good populations.

This criterion implies that one would prefer the selection with probability  $\gamma$  of all good populations and with probability  $1 - \gamma$  of none of them to the selection with probability  $\gamma - \epsilon$  of all and with probability  $1 - \gamma + \epsilon$  of all but one of the good populations. The criterion would thus seem to be appropriate only in rare cases.

As a measure of the performance of a procedure with respect to false positives we shall take either

(i) the expected number of false positives

or

(ii) the expected proportion of false positives, that is, the quantity (i) divided by the total number of negatives.

As a generic notation for any one of the quantities (a)-(e), all of which depend on the parameter point  $\theta$  and on the particular selection procedure  $\delta$  under investigation, we shall use  $S(\theta, \delta)$ . Here it is to be understood that  $S$  is defined only for the set  $\Omega'$  of those parameter-points for which at least one of the populations is positive.

Similarly, we shall let  $R(\theta, \delta)$  denote the quantity (i) or (ii). With these definitions of  $R$  and  $S$ , it is desirable to have  $S(\theta, \delta)$  as large and  $R(\theta, \delta)$  as small as possible. Specifically, we shall consider the problem of determining a procedure for which, subject to

$$(4) \quad \inf_{\theta \in \Omega'} S(\theta, \delta) \geq \gamma$$

we have

$$(5) \quad \sup_{\theta \in \Omega} R(\theta, \delta) = \min,$$

(where  $\Omega$  denotes the whole parameter space), or the dual problem in which  $\inf S(\theta, \delta)$  is maximized subject to an upper bound on  $\sup R(\theta, \delta)$ .

Which of the various formulations is most appropriate, depends of course on the particular circumstances of each problem. In the absence of such more specific

<sup>3</sup> A complete sequential procedure for dealing with this problem was proposed by Stein [28]. For more recent work on such sequential procedures, see [29].

considerations, it seems perhaps most reasonable to control the minimum value of either (b) or (d). Subject to this condition one might, since each false positive provides a nuisance disturbance, wish to minimize the maximum value of (i).

It is of interest to note that condition (4) with  $S$  given by (b), that is,

$$(6) \quad \inf_{\theta \in \Omega'} [\text{expected proportion of true positives}] \geq \gamma,$$

implies

$$(7) \quad \inf_{\theta \in \Omega'} P\{\text{at least one true positive}\} \geq \gamma.$$

This follows from the fact that the left-hand side of (7) always exceeds that of (6). To see this, denote by  $A_i$  the event of including the  $i$ th population in the selected group, let  $I$  denote the set of indices  $i$  for which the  $i$ th population is positive, and let  $k$  be the number of elements of  $I$ . Then

$$P(\bigcup_{i \in I} A_i) \geq \max_{i \in I} P(A_i) \geq \sum_{i \in I} P(A_i)/k$$

as was to be proved.

The other proposed condition, (4) with  $S$  given by (d), that is,

$$(8) \quad P\{\text{best population is in selected group}\} \geq \gamma \quad \text{for all } \theta \in \Omega',$$

can be given the following interpretation. Let  $\Sigma$  denote the set of selected indices. Then  $\Sigma$  constitutes a confidence set for the index  $i$  corresponding to the best population, provided attention is restricted to the parameter set  $\Omega'$ .

We shall prove in the next sections, for certain families of distributions, that the solution with any of the formulations (a)–(d) combined with either (i) or (ii) is given by (3) of section 1 but that this is not true for formulation (e).

**4. A minimax solution.** If attention is restricted to nonrandomized procedures, as can always be done by enlarging the sample space, a selection procedure is a partition of the sample space into the sets  $D_{i_1, \dots, i_k}$  of those sample points for which the selected group consists of the populations with subscripts  $i_1, \dots, i_k$  and no others. To these must be added the set  $D_0$  for which none of the populations is selected. If the number of available populations is  $a$ , the number of sets  $D$  is  $2^a$  since each subscript may or may not occur with all combinations of the remaining subscripts.

Fortunately, selection procedures possess an equivalent, and for most purposes much simpler, representation. Let  $E_i$  be the set of sample points for which the  $i$ th population is included in the selected group. Then each of the two systems of sets  $\{D\}$  and  $\{E\}$  is uniquely determined by the other. In fact,  $E_i$  is the union of all those sets  $D$  which have  $i$  as one of their subscripts. Conversely,

$$D_{i_1, \dots, i_k} = E_{i_1} \cap \dots \cap E_{i_k} \cap \bar{E}_{j_1} \cap \dots \cap \bar{E}_{j_{a-k}}$$

where  $j_1, \dots, j_{a-k}$  are the subscripts different from  $i_1, \dots, i_k$  and  $\bar{E}$  denotes the complement of  $E$ . Instead of working with the sets  $E_i$  in the enlarged sample space, it is now more convenient to return to the possibility of randomization.

Each  $E_i$  is then represented by a function  $\psi_i$  defined over the sample space and taking on values between 0 and 1, where  $\psi_i(x)$  denotes the probability with which the  $i$ th population is included in the selected group. A selection procedure is characterized by the vector  $\psi = (\psi_1, \dots, \psi_a)$ .<sup>4</sup>

According to the formulation of the preceding section we are concerned with a minimax problem subject to side conditions. This type of problem was investigated by Blyth [31] but his conditions do not apply to the cases to be considered here. The following lemma is an immediate extension of the standard method of characterizing minimax solutions as Bayes solutions corresponding to a least favorable *a priori* distribution. As in its more usual form, it is essentially an application of the Lagrange method of undetermined multipliers.

LEMMA 1. Let  $\mathfrak{B}$  be a  $\sigma$ -field of subsets of the parameter space  $\Omega$  and let  $\lambda$  and  $\mu$  be probability distributions over  $(\Omega, \mathfrak{B})$ . Let  $A, B$  be two positive constants and let  $\delta_0$  maximize the integral

$$(9) \quad B \int S(\theta, \delta) d\mu(\theta) - A \int R(\theta, \delta) d\lambda(\theta).$$

Then  $\delta_0$  minimizes  $\sup R(\theta, \delta)$  subject to

$$(10) \quad \inf S(\theta, \delta) \geq \gamma$$

provided

$$(11) \quad \int R(\theta, \delta_0) d\lambda(\theta) = \sup R(\theta, \delta_0)$$

and

$$(12) \quad \int S(\theta, \delta_0) d\mu(\theta) = \inf S(\theta, \delta_0) = \gamma.$$

If  $\delta_0$  is the unique procedure maximizing (9), it is also the unique solution of the restricted minimax problem.

PROOF. Let  $\delta$  be any procedure satisfying (10). Then

$$\begin{aligned} B \int S(\theta, \delta) d\mu(\theta) - A \int R(\theta, \delta) d\lambda(\theta) \\ \leq B \int S(\theta, \delta_0) d\mu(\theta) - A \int R(\theta, \delta_0) d\lambda(\theta) = B\gamma - A \sup R(\theta, \delta_0). \end{aligned}$$

<sup>4</sup> This representation was first utilized in a slightly more special form by Robbins [23]. A generalization was given by the author in Theorem 1 of [30]. I am grateful to Professor L. LeCam for pointing out an error in the generalization of Theorem 1 to randomized procedures. The displayed equivalence formulae at the top of p. 6 of [30] are not correct. However, the equivalence theorem itself remains correct even when randomization is permitted. This can be seen as above, by representing a randomized procedure as a nonrandomized procedure in an enlarged sample space and applying Theorem 1.

Since  $\delta$  satisfies (10), it follows that

$$\sup R(\theta, \delta_0) \leq \int R(\theta, \delta) d\lambda(\theta) \leq \sup R(\theta, \delta)$$

as was to be proved.

If condition (10) is to hold only when the sup is taken over a subset  $\Omega'$  of  $\Omega$ ,  $\lambda$  must be a distribution over  $\Omega'$  but no other changes are necessary.

As is usually the case with minimax problems, the more difficult part of the solution is not the maximization of (9) but the determination of an appropriate  $\lambda$  and  $\mu$ . In this connection, the following standard devices are helpful.

1. Condition (11) implies that  $\lambda$  assigns probability one to the set  $\omega$  of parameter points  $\theta$  for which

$$(13) \quad R(\theta, \delta_0) = \sup_{\theta'} R(\theta', \delta_0).$$

Similarly,  $\mu$  must assign probability one to the set for which

$$S(\theta, \delta_0) = \inf_{\theta'} S(\theta', \delta_0).$$

2. The pair  $(\lambda, \mu)$  is least favorable in the sense that it minimizes the maximum (with respect to  $\delta$ ) value of (9).

3. If the problem exhibits any symmetries, it pays to look for distributions  $\lambda, \mu$  possessing the corresponding symmetries.

We shall in the following consider procedures, which determine the selection or nonselection of the  $i$ th population on the basis of real-valued statistics  $T_i$ , and in particular we shall prove certain minimax properties for procedures of the type

$$(14) \quad \psi_i = 1, \lambda_i, 0 \quad \text{as } T_i >, =, < C_i.$$

For this purpose it is convenient first to state the following lemma, the proof of which is immediate.

LEMMA 2. Suppose that the distribution of  $T_i$  depends only on  $\theta_i$  and is stochastically increasing in  $\theta_i$ . Let  $\delta = (\psi_1, \dots, \psi_a)$  be any procedure satisfying (14) and let  $I$  be the set of subscript  $i$  for which

$$E_{\theta_0 + \Delta} \psi_i = \min_{j=1, \dots, a} E_{\theta_0 + \Delta} \psi_j.$$

Then, if  $S$  is given by one of the quantities (a), (c) or (d) of Section 3,  $\inf_{\Omega'} S(\theta, \delta)$  is attained at all points  $\theta$  such that for some  $i \in I$

$$\theta_i = \theta_0 + \Delta \quad \text{and} \quad \theta_j < \theta_0 + \Delta \quad \text{for all } j \neq i.$$

If  $S$  is given by (b),  $\inf_{\Omega'} S(\theta, \delta)$  is attained at all points  $\theta$  such that for some subset  $\{i_1, \dots, i_k\}$  of  $I$

$$\theta_{i_1} = \dots = \theta_{i_k} = \theta_0 + \Delta \quad \text{and} \quad \theta_j < \theta_0 + \Delta \quad \text{for the remaining } \theta_j\text{'s}.$$

If  $R$  is given by (i) or (ii), then  $\sup_{\Omega} R(\theta, \delta)$  is attained at the point

$$\theta^{(0)} = (\theta_0, \dots, \theta_0).$$

We note also that if in addition to the assumptions made above, the joint distribution of  $(T_1, \dots, T_a)$  is stochastically increasing in  $(\theta_1, \dots, \theta_a)$ , and if  $S$  is given by (e) of Section 3, then  $\inf_{\theta} S(\theta, \delta)$  is attained at the point  $(\theta_0 + \Delta, \dots, \theta_0 + \Delta)$ .

The minimax solution can now be obtained under the following conditions.

**THEOREM.** Let the probability density of  $X$  be denoted by  $p_0$  when  $\theta_1 = \dots = \theta_a = \theta_0$ , and by  $p_i$  when  $\theta_i = \theta_0 + \Delta$  and the parameters  $\theta_j$  for  $j \neq i$  have a common value  $\theta' \leq \theta_0 + \Delta$  determined so that the conditions below are satisfied. Suppose that  $p_i(x)/p_0(x)$  is a nondecreasing function of a real-valued statistic  $T_i$ , that the distribution of  $T_i$  depends only on  $\theta_i$ , is stochastically increasing in  $\theta_i$ , and is independent of  $i$ . Then the procedure  $\delta_0$  satisfying (4) and (5) with  $S$  equal to any one of the quantities (a)–(d) and  $R$  defined by (i) or (ii) of the preceding section, is given by

$$(15) \quad \psi_i = 1, \lambda_0, 0 \quad \text{as } T_i >, =, < C,$$

where  $\lambda_0$  and  $C$  are determined by

$$(16) \quad E_{\theta_0 + \Delta} \psi_i = \gamma.$$

The solution of the dual problem in which (4) is replaced by

$$(17) \quad \sup R(\theta, \delta) \leq \gamma'$$

is also given by (15), with  $\lambda$  and  $C$  now determined by

$$(18) \quad R(\theta^0, \delta_0) = \gamma'$$

where  $\theta^0 = (\theta_0, \dots, \theta_0)$ .

**PROOF.**

1. Let  $\mu$  be the distribution which assigns probability one to the point  $(\theta_0, \dots, \theta_0)$  and  $\lambda$  the distribution which assigns probability  $1/a$  to the points  $\theta^{(i)}$  given by

$$(19) \quad \theta_i = \theta_0 + \Delta, \quad \theta_j = \theta' < \theta_0 + \Delta \quad \text{for } j \neq i.$$

Then it follows from Lemma 2 that  $\delta_0$  satisfies conditions (11) and (12).

2. For the distributions  $\lambda$  and  $\mu$  specified in 1, and with  $p_i$  denoting the probability density of  $X$  when  $\theta = \theta^{(i)}$ , we have

$$(20) \quad R(\theta^0, \delta) = \int (\psi_1 + \dots + \psi_a) p_0; \quad S(\theta^{(i)}, \delta) = \int \psi_i p_i$$

and hence (9) reduces to

$$(21) \quad \frac{B}{a} \sum_{i=1}^a \int \psi_i p_i - A \sum_{i=1}^a \int \psi_i p_0 = \int \sum \psi_i \left( \frac{B}{a} p_i - A p_0 \right).$$

Since  $0 \leq \psi_i \leq 1$ , (21) is maximized by putting  $\psi_i = 0$  or 1 as

$$(B/a)p_i < \text{ or } > Ap_0,$$

and hence as  $T_i$  is  $<$  or  $>$   $C$ , as was to be proved.

We note that it is actually not necessary for  $p_i(x)/p_0(x)$  to be an increasing function of  $T_i$ , but only that there exist a constant  $k$  such that the regions  $T_i > C$  and  $T_i < C$  (for the particular value  $C$  determined by the side conditions) are equivalent to the regions  $p_i(x)/p_0(x) > k$  and  $< k$  respectively.

This theorem provides the basis for determining the sample size necessary to control the risks  $R$  and  $S$  at any desired levels. For suppose that we wish the selection procedure to satisfy

$$R(\theta, \delta) \leq \gamma' \quad \text{for all } \theta \in \Omega$$

and

$$S(\theta, \delta) \geq \gamma \quad \text{for all } \theta \in \Omega'.$$

Then for the smallest sample size (possibly randomized) which constitutes a solution to this problem, the associated procedure  $\delta_0$  minimizes  $\sup R(\theta, \delta)$  subject to (4). If the conditions of the theorem are satisfied,  $\delta_0$  is therefore given by (15) and hence satisfies (if we assume for simplicity that it is nonrandomized)

$$\sup R(\theta, \delta_0) = R(\theta_0, \delta_0) = \begin{cases} aP_{\theta_0}(T_i \geq C) & \text{if } R \text{ is given by (i)} \\ P_{\theta_0}(T_i \geq C) & \text{if } R \text{ is given by (ii)}. \end{cases}$$

It further satisfies the condition

$$\inf S(\theta, \delta_0) = P_{\theta_0+\Delta}(T_i \geq C).$$

If we let

$$\gamma^* = \begin{cases} \gamma'/a & \text{when } R \text{ is given by (i)} \\ \gamma' & \text{when } R \text{ is given by (ii)}, \end{cases}$$

the sample size is therefore determined by the conditions

$$P_{\theta_0}(T_i \geq C) \leq \gamma^*, \quad P_{\theta_0+\Delta}(T_i \geq C) \geq \gamma.$$

These are exactly the conditions appropriate for testing the hypothesis  $\theta_i = \theta_0$  against the alternative  $\theta_i = \theta_0 + \Delta$  if we wish to have significance level  $\gamma^*$  and power at least  $\gamma$ . In all particular cases considered in the following section, the sample size determination therefore reduces to a problem whose solution is known from the corresponding problem of hypothesis testing.

**5. Families with monotone likelihood ratio.** The theorem of the preceding section applies directly to the case of independent samples  $X_{i1}, \dots, X_{in}$  from populations with probability density  $f_{\theta_i}$  depending only on the real-valued parameter  $\theta_i$  with respect to which we wish to select, if there exists a sufficient statistic  $T_i$  for  $(X_{i1}, \dots, X_{in})$  with monotone likelihood ratio. Let the probability density of  $T_i$  be  $g_{\theta_i}$ , and take  $\theta_0$  for the value  $\theta'$  of the theorem. Then

$$\frac{p_i(x)}{p_0(x)} = \frac{g_{\theta_0+\Delta}(t_i)}{g_{\theta_0}(t_i)} \prod_{j \neq i} \frac{g_{\theta_0}(t_j)}{g_{\theta_0}(t_j)} = \frac{g_{\theta_0+\Delta}(t_i)}{g_{\theta_0}(t_i)},$$

which is nondecreasing in  $t_i$ , and it follows that the minimax procedure is given by (15).



In particular, if  $n = 1$  and the probability densities  $f_{\theta_i}(x_i)$  have monotone likelihood ratio in  $x_i$ , the result holds with  $T_i = X_i$  so that the procedure is given by

$$(22) \quad \psi_i = 1, \lambda_0, 0 \quad \text{as } X_i >, =, < C.$$

Examples of this are the case in which the random variables  $X_i$  are independently distributed with binomial distributions  $b(p_i, m)$ , with Poisson distributions  $P(\tau_i)$ , or more generally with distributions having densities of the form

$$C(\theta_i)e^{\theta_i x_i}h(x_i),$$

that is, belonging to an exponential family.

As another example, consider samples  $(X_{i1}, \dots, X_{in})$  from normal distributions  $N(\xi_i, \sigma_i^2)$  and suppose we wish to select the populations with small variances. Attention may be restricted to the sufficient statistic  $\bar{X}_1, \dots, \bar{X}_a$  and  $S_1^2, \dots, S_a^2$  where

$$\bar{X}_i = \sum_{j=1}^n X_{ij}/n; \quad S_i^2 = \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2.$$

Since the problem remains invariant<sup>5</sup> under addition of arbitrary constants  $c_i$  to  $\bar{X}_i$ , it follows from a trivial extension of the Hunt-Stein theorem ([32], p. 336) that there exists a minimax solution depending only on the variables  $S_1^2, \dots, S_a^2$ . To these variables the theorem is now applicable with  $n = 1$ , and shows that the minimax procedure consists in selecting the populations for which  $S_i^2 \leq C$ .

Suppose that instead it is desired to select the populations for which the parameters  $\theta_i = \xi_i/\sigma_i$  are sufficiently large. This time the problem remains invariant under multiplication of  $\bar{X}_i$  by any positive constant  $c_i$  and of  $S_i^2$  by  $c_i^2$ . There exists therefore a minimax solution depending only on the variables  $T_i = \bar{X}_i/S_i$ . Since the  $T_i$  have noncentral  $t$ -distributions which possess monotone likelihood ratio, it follows that the minimax procedure consists in selecting the populations for which  $\bar{X}_i/S_i \geq C$ .

If in this last problem the variances  $\sigma_i^2$  are assumed to be independent of  $i$ , with the common variance  $\sigma^2$  still being unknown, the problem (of selecting for  $\xi_i/\sigma$  or  $\xi_i$ ) surprisingly is much less simple since the  $T_i$  are then dependent. We shall consider this case in Section 7.

We conclude the present section by showing that in all the problems considered above, if criterion (e) of Section 1 is used instead of one of the criteria (a)-(d), the minimax solution is no longer given by (15). The argument is sufficiently clearly indicated by considering the case  $a = 2$ . Suppose therefore that  $X_1, X_2$  are independently distributed with densities  $f_{\theta_1}(x)$  and  $f_{\theta_2}(x)$  which have monotone likelihood ratio. We shall also assume that the associated cumulative distribution functions are continuous; if the region of positive density is

<sup>5</sup> In the same strong sense as in the theory of hypothesis testing that is without performing any transformations of the decision space.

independent of  $\theta$ , this can always be achieved by adjoining a uniformly distributed random variable.

Subject to

$$(23) \quad \sup [\text{Expected number of false positives}] \leq \gamma'$$

we wish to maximize the minimum probability of including in the selected group all good populations (if there are any). The procedure of the preceding section includes the  $i$ th population when  $X_i \geq C$  where  $P_{\theta_0}\{X_i \geq C\} = \gamma'/2$ . Then the maximum expected number of false positives is exactly  $\gamma'$  and, if

$$\beta = P_{\theta_0+\Delta}(X_i \geq C),$$

the minimum probability of including all good populations is  $\beta^2$ .

Consider now the following alternative procedure:

Include  $\theta_1$  in the selected group

if  $X_1 \geq C$  and if  $X_2 < C - \epsilon$  or  $X_2 \geq C$ .

Include  $\theta_2$  in the selected group

if  $X_2 \geq C$  and if  $X_1 < C - \epsilon$  or  $X_1 \geq C$ .

In addition include both  $\theta_1$  and  $\theta_2$  if  $C - \epsilon \leq X_1, X_2 < C$ .

For  $\epsilon$  sufficiently small, it is then easily checked that the probability of including  $\theta_i$  in the selected group when  $\theta_i \leq \theta_0$  is still  $\leq \gamma'/2$  so that the procedure continues to satisfy (23). The probability of including in the selected group all good populations if only one of the  $\theta$ 's is  $\geq \theta_0 + \Delta$  is now less than its previous value  $\beta$  but by continuity is  $> \beta^2$  for  $\epsilon$  sufficiently small. On the other hand, the probability of including all good populations when both  $\theta_1$  and  $\theta_2$  are  $\geq \theta_0 + \Delta$  is now clearly  $> \beta^2$  since the set of sample points for which both populations are included has been increased by the set  $C - \epsilon \leq X_1, X_2 < C$ . Hence for  $\epsilon$  sufficiently small, the minimum probability of including all good populations is now  $> \beta^2$  as was to be proved.

**6. Unequal sample sizes.** The case of unequal sample sizes requires a slight generalization of the theorem of Section 4.

**THEOREM.** Let  $T_i$  ( $i = 1, \dots, a$ ) be a real valued statistic whose distribution depends only on  $\theta_i$  and is stochastically increasing in  $\theta_i$ . Let  $\delta_0 = (\psi_1, \dots, \psi_a)$  be the selection procedure defined by

$$(24) \quad \psi_i = 1, \lambda_i, 0 \quad \text{as } T_i >, =, < C_i$$

where  $\lambda_i, C_i$  are determined by (16). Let  $p_i(x)$  and  $p_0(x)$  be defined as in Section 4 and suppose that there exist constants  $k_i$  such that

$$p_i(x)/p_0(x) >, =, < k_i \quad \text{as } T_i >, =, < C_i.$$

Then, if  $S$  is equal to one of the quantities (a)-(d) and  $R$  is defined by (i) or (ii),  $\delta_0$  minimizes  $\sup_{\theta \in \Omega} R(\theta, \delta)$  subject to (4). If instead of by (16), the constants  $\lambda_i, C_i$  are determined by (18) and

$$(25) \quad E_{\theta_0+\Delta} \psi_i \text{ is independent of } i,$$

$\delta_0$  maximizes  $\inf_{\theta \in \Omega} S(\theta, \delta)$  subject to (17).

Thus in both cases, the critical functions  $\psi_i$ , which can be interpreted as tests of the hypotheses  $H_i: \theta_i \leq \theta_0$  against the alternatives  $K_i: \theta_i \geq \theta_0 + \Delta$ , are not determined to give a constant (independent of  $i$ ) significance level but instead so that the minimum power against  $K_i$  is constant.

The reason for this is clear: the probability of the  $i$ th population giving rise to a false positive takes on its maximum value at the same point  $\theta^{(0)}$  for all  $i$ ; hence the contributions of the various populations to the expected number or proportion of false positives are combined and do not have to be controlled individually. On the other hand, the probability of the  $i$ th population resulting in a true positive takes on its minimum value at a different point  $\theta^{(i)}$  for each  $i$ . For the minimum of these minima to be a maximum, they have to be equal.

The lack of symmetry shown by the solution is of course a consequence of the asymmetric formulation of Section 3, where the consideration was shifted from false negatives to true positives. However, this asymmetry is not artificial but only reflects a corresponding asymmetry of the problem.

The proof of the result for unequal sample sizes parallels that of the special case when the sample sizes are equal. Let  $\delta_0$  be the procedure determined by (24), (25) and (16) or (18). Let the distribution  $\mu$  be defined as before but let  $\lambda$  assign to the point  $\theta^{(i)}$ , instead of  $1/a$ , a probability  $\pi_i$  to be determined later. The quantity (21) then becomes

$$(26) \quad \int \sum \psi_i (B\pi_i p_i - Ap_0)$$

which is maximized by putting  $\psi_i = 0$  or  $1$  as

$$p_i/p_0 < \text{or} > A/B\pi_i.$$

Hence if  $B/A$  and  $\pi_i$  are determined so that  $A/B\pi_i$  is equal for each  $i$  to the constant  $k_i$  defining  $\delta_0$ , it follows that  $\delta_0$  has the desired minimax property.

**7. Normal populations with common unknown variance.** Let  $X_{ij}$  ( $j = 1, \dots, n_i$ ;  $i = 1, \dots, a$ ) be normally distributed as  $N(\xi_i, \sigma^2)$ , and suppose we wish to select the populations with large values of  $\xi_i/\sigma$ . More specifically, we shall consider a population as negative if  $\xi_i/\sigma \leq 0$  and as positive if  $\xi_i/\sigma \geq \Delta$ . A set of sufficient statistics is given by the means  $\bar{X}_i$  together with

$$S^2 = \sum \sum (X_{ij} - \bar{X}_i)^2.$$

Since the problem remains invariant under multiplication of each  $\bar{X}_i$  by the same positive constant  $c$  and of  $S^2$  by  $c^2$ , there exists a minimax procedure depending only on the variables  $Y_i = \bar{X}_i/S$ . If  $\theta_i = \xi_i/\sigma$ , the joint density of the  $Y$ 's, is up to a constant,

$$(27) \quad p(y_1, \dots, y_a) = (1 + \sum n_i y_i^2)^{-a-f/2} \cdot \int_0^\infty w^{2a+f-1} \exp[-w^2/2 + w \sum n_i \theta_i y_i / \sqrt{1 + \sum n_i y_i^2}] dw,$$

where

$$f = \sum_{i=1}^a (n_i - 1).$$

From a consideration of the individual hypotheses  $H_i: \theta_i \leq 0$ , it appears natural to include the  $i$ th population in the selected group if

$$(28) \quad \bar{X}_i/S \geq C_i.$$

For this procedure it follows as before that the expected number of proportion of false positives takes on its maximum value at  $\theta^{(0)} = (0, \dots, 0)$  while the quantities (a)-(d) take on their minimum value at (among other points) the points  $\theta^{(i)} = (0, \dots, 0, \Delta, 0, \dots, 0)$ .

If the joint density of the  $Y$ 's at  $\theta^{(0)}$  and  $\theta^{(i)}$  is denoted by  $p_0$  and  $p_i$ , it is seen from (27) that  $p_i/p_0$  is an increasing function of

$$\frac{\Delta Y_i}{\sqrt{1 + \sum n_j Y_j^2}} = \frac{\Delta \bar{X}_i}{\sqrt{S^2 + \sum n_j \bar{X}_j^2}}.$$

The selection of the  $i$ th population when  $p_i/p_0$  is sufficiently large thus leads to selecting the populations for which

$$(29) \quad \frac{\bar{X}_i}{\sqrt{S^2 + \sum n_j \bar{X}_j^2}} > C_i.$$

This corresponds to the solution proposed by Paulson [15] and Pfanagl [18] for the associated slippage problem in which the standard is replaced by a control.<sup>6</sup> It is however not a solution to the present problem. For as  $\theta_j \rightarrow -\infty$  for some  $j \neq i$ , the probability of the inequality (29) tends to zero, and so therefore does the minimum value of each of the quantities (a)-(d).

Sometimes it is not unreasonable to assume *a priori* that

$$(30) \quad \theta_i \geq 0 \quad \text{for all } i.$$

If for example we wish to select among a number of possible enrichments of a substandard diet, we may be willing to assume that the effect of each, if any, is beneficial. While under this assumption, for most significance levels and sample sizes, the performance of the procedure (29) is no longer as drastic as before, the procedure is nevertheless still not a minimax solution. This is easily seen from the fact that the probability of the inequality (29), subject to  $\theta_i = \Delta$  and  $0 \leq \theta_j \leq \Delta$  for all  $j \neq i$  takes on its minimum value not at  $\theta^{(i)}$  but instead at the point  $\theta_1 = \dots = \theta_a$ . The proof is similar (but simpler) to the one given below.

The question arises whether the intuitive procedure (28) is, as one might expect, a minimax solution. We shall prove in the next section that this is not the case, under the *a priori* assumption (30). It seems likely that the situation is the

<sup>6</sup> A slippage procedure, corresponding to (28) was proposed by Paulson in [9]; the complete procedure (28), with the standard replaced by a control, is discussed by Dunnett [33].

same even without this assumption. However, as we shall now show, (28) is then at least approximately minimax in the sense that subject to (4), the maximum expected number of false positives can at least be improved only slightly over its value for (28). We shall show this with the constant  $C_i$  determined by (25).

Condition (4) implies that for all  $i$ ,

$$(31) \quad E\psi_i \geq \gamma \quad \text{when} \quad \theta_i \geq \Delta \quad \text{and} \quad \theta_j < \Delta \quad \text{for } j \neq i.$$

Since the procedure (28), which for the moment we shall denote by  $\psi^* = (\psi_1^*, \dots, \psi_a^*)$ , attains the maximum expected number of false positives when  $\theta_1 = \dots = \theta_a = 0$ , the minimax value of this quantity can be no higher than

$$E_{0,\dots,0}(\psi_1^* + \dots + \psi_a^*) = \alpha_1 + \dots + \alpha_a$$

where

$$\alpha_i = P_0\{\bar{X}_i/S > C_i\},$$

$C_i$  being determined by  $E_{\Delta}\psi_i^* = \gamma$ .

On the other hand, this minimax value cannot be much lower than  $\sum \alpha_i$ . For consider any procedure  $\psi = (\psi_1, \dots, \psi_a)$  satisfying (31). Its maximum expected number of false positives is greater than or equal to  $E_{0,\dots,0}(\psi_1 + \dots + \psi_a)$ . Consider now the problem of minimizing  $E_{0,\dots,0}(\psi_1 + \dots + \psi_a)$  subject to (31). If we restrict (31) to the parameter values  $\theta_i \geq \Delta$  and  $\theta_j = 0$  for  $j \neq i$ , the solution becomes

$$\psi_i = 1 \quad \text{when} \quad \bar{X}_i / \sqrt{S^2 + \sum_{j \neq i} n_j \bar{X}_j^2} > C'_i.$$

Thus  $\sum \alpha'_i$  where  $\alpha'_i = P_{0,\dots,0}\{\bar{X}_i / \sqrt{S^2 + \sum_{j \neq i} n_j \bar{X}_j^2} > C'_i\}$  constitutes a lower bound for the sought for minimax value. However, for typical values of the sample sizes,  $\alpha'_i$  will be only slightly lower than  $\alpha_i$ , the only difference being the  $a - 1$  added degrees of freedom in the denominator of the  $t$ -statistic, which now has  $\sum n_i - 1$  degrees of freedom instead of the  $\sum (n_i - 1)$  in the case of (28).

The same argument shows that (28) approximately minimizes the maximum expected proportion of false positives.

**8. A counterexample.** We shall now construct, for the case of equal sample sizes, a procedure satisfying (23) and with larger minimum value of (a)-(d) than (28). To this end we shall first prove the existence of points  $(y_1, \dots, y_a)$  and  $(y'_1, \dots, y'_a)$  with  $y_1 < C < y'_1$  and such that

$$\frac{p_{\Delta, \theta_2, \dots, \theta_a}(y_1, \dots, y_a)}{p_{0,0,\dots,0}(y_1, \dots, y_a)} > \frac{p_{\Delta, \theta_2, \dots, \theta_a}(y'_1, \dots, y'_a)}{p_{0,0,\dots,0}(y'_1, \dots, y'_a)}$$

for all  $0 \leq \theta_2, \dots, \theta_a \leq \Delta$ . It is seen from (27), that the probability ratio  $p_{\Delta, \theta_2, \dots, \theta_a} / p_{0,0,\dots,0}$  is an increasing function of

$$(32) \quad (\Delta y_1 + \sum_{j=2}^a \theta_j y_j) / \sqrt{1 + n \sum_{j=1}^a y_j^2}.$$

Hence it is enough to construct the points  $y$  and  $y'$  in such a way that the expression (32) (where without loss of generality we can put  $n = 1$ ) exceeds the corresponding expression when  $y$  is replaced by  $y'$ . Putting  $y_2 = \dots = y_a$ ,  $y'_2 = \dots = y'_a$ , letting  $\rho = (\theta_2 + \dots + \theta_a)/\Delta$  and writing  $k$  for  $a - 1$ , we must find pairs  $(y_1, y_2)$ ,  $(y'_1, y'_2)$  with  $y_1 < C < y'_1$  and such that

$$\frac{y_1 + \rho y_2}{\sqrt{1 + y_1^2 + (a-1)y_2^2}} > \frac{y'_1 + \rho y'_2}{\sqrt{1 + y'^2_1 + (a-1)y'^2_2}} \quad \text{for all } 0 \leq \rho \leq k.$$

Since all coordinates will be chosen to be nonnegative, the inequality can be squared and on collection of powers of  $\rho$  becomes

$$f(\rho) = a_0 \rho^2 + 2a_1 \rho + a_2 > 0$$

with

$$a_0 = y_2^2(1 + y'^2_2) - y'^2_2(1 + y_1^2)$$

$$a_1 = y_1 y_2 [1 + y'^2_2 + (a-1)y'^2_2] - y'_1 y'_2 [1 + y_1^2 + (a-1)y_2^2]$$

$$a_2 = y_1^2 [1 + (a-1)y'^2_2] - y'^2_1 [1 + (a-1)y_2^2].$$

A sufficient set of conditions for  $f(\rho)$  to be positive for all  $0 \leq \rho \leq k$  is  $a_0 < 0$ ,  $f(0) > 0$ ,  $f(k) > 0$ , and hence

$$a_0 < 0, \quad a_2 > 0, \quad a_0 k^2 + 2a_1 k + a_2 > 0.$$

For any fixed  $y_1$ ,  $y'_1$  and  $y_2$ , the first two of these conditions are satisfied if  $y'_2$  is sufficiently large. The coefficient of  $y'^2_2$  in  $a_0 k^2 + 2a_1 k + a_2$  is  $-k^2(1 + y_1^2) + 2ky_1 y_2(a-1) + (a-1)y_1^2$ . If  $y_2$  is chosen large enough so that this coefficient is positive, the third condition is also satisfied for  $y'_2$  sufficiently large, and this completes the proof. The two points constructed in this way will be denoted by  $y^0 = (y^0_1, \dots, y^0_a)$  and  $y^{0'} = (y^{0'}_1, \dots, y^{0'}_a)$ . We note for later use that the points can be chosen in such a way that  $C < y_j, y'_j$  for all  $j > 1$ .

Let  $R$  and  $R'$  denote two spheres with centers at  $y^0$  and  $y^{0'}$ , and radii determined so that

$$(33) \quad P(R) = P(R') \quad \text{when } \theta_1 = \dots = \theta_a = 0.$$

In addition, the spheres are to be sufficiently small so that

$$(34) \quad \frac{y_1 + \sum_{j=2}^a \rho_j y_j}{\sqrt{1 + \sum_{j=2}^a y_j^2}} > \frac{y'_1 + \sum_{j=2}^a \rho_j y'_j}{\sqrt{1 + \sum_{j=2}^a y'^2_j}} \quad \text{for all } 0 \leq \rho_j \leq 1,$$

that  $y_1 < C < y'_1$  for all  $y \in R$  and  $y' \in R'$  and that further conditions are satisfied which will be specified later.

Consider now the following modification of (28):

the 1st population is selected if the sample point satisfies

$$(35) \quad (y \in R) \quad \text{or} \quad (y_1 > C \quad \text{and} \quad y \notin R'),$$

and the rule for selecting the other populations is defined by symmetry.



By (35), the expected number of false positives, when  $\theta_1 = \dots = \theta_a = 0$ , is the same for the modified procedure as for (28). Since under assumption (30) the  $i$ th population can give rise to a false positive only when  $\theta_i = 0$ , it follows that the modified procedure satisfies (23) if  $R$  and  $R'$  are sufficiently small.

Consider on the other hand one of the criteria (a)-(d), for example the probability of including the best population in the selected group when its  $\theta$ -value is  $\geq \Delta$ . Under (28) this attains its minimum value at the points whose coordinates for some  $i$  satisfy  $\theta_i = \Delta$  and  $0 \leq \theta_j < \Delta$  for  $j \neq i$ . By (34), the probability of including the best population is larger at all these points under the modified procedure than under (28).

In order to prove that also the minimum probability of including the best population has been increased by the modification, it is sufficient to show that with the modified procedure this minimum probability is still attained at points satisfying  $\theta_i = \Delta$ ,  $0 \leq \theta_j < \Delta$  for  $j \neq i$ . This follows easily if we can show that for the modified procedure the probability of including the  $i$ th population is an increasing function of  $\theta_i$  for fixed values of the other  $\theta$ 's. This result finally is an immediate consequence of the following two facts.

1. The partial derivative

$$\frac{\partial}{\partial \theta_i} P_{\theta_i} \{ \bar{X}_i / S > C \}$$

is positive for all  $\theta_i$  and is bounded away from 0 in any finite interval  $a \leq \theta_i \leq b$ .

2. As the radius of the sphere  $R$  (and hence also of  $R'$ ) tends to zero, the derivatives

$$\frac{\partial}{\partial \theta_i} P_{\theta_1, \dots, \theta_a}(R) \quad \text{and} \quad \frac{\partial}{\partial \theta_i} P_{\theta_1, \dots, \theta_a}(R')$$

tend to zero.

PROOF.

1. Putting  $\sigma = 1$  so that  $\xi_i = \theta_i$ , the derivative is equal to

$$\frac{\partial}{\partial \theta_i} \int_0^\infty P_{\theta_i} \{ \bar{X}_i > C\varepsilon \} dP(s).$$

We can differentiate under the integral sign and the derivative of the integrand is known to be positive. (See for example [32], p. 114. Problem 18.) Since this derivative is a continuous function of  $\theta_i$ , it is bounded away from zero in any finite interval.

2. Writing  $P(R) = \int_R P_{\theta_1, \dots, \theta_a}(y) dy$ , the differentiation can be carried out under the integral sign. Since for  $(\theta_1, \dots, \theta_a)$  in any finite interval, the integrand is uniformly bounded, the result follows.

Exactly the same argument applies if criterion (d) is replaced by (a) or (c). However, with (b) the difficulty arises that the expected proportion of true positives takes on its minimum value at all points which for some  $1 \leq i_1 < \dots < i_k \leq a$ ,  $1 \leq k \leq a$ , satisfy

$$\theta_{i_1} = \dots = \theta_{i_k} = \Delta; \quad 0 \leq \theta_j < \Delta \quad \text{for all } j \neq i_1, \dots, i_k.$$

To obtain an increase at all these points, we note that at the beginning of the section we proved the existence of points  $y = (y_1, \dots, y_a)$  and  $y' = (y'_1, \dots, y'_a)$  with

$$y_1 < C < y'_1 \quad \text{and} \quad C < y_j, y'_j \quad \text{for all } j \neq 1$$

and such that (34) holds.

Let  $R_1$  and  $R'_1$  denote the spheres previously denoted by  $R$  and  $R'$ , let  $R_i$  and  $R'_i$  be defined by symmetry. The modified procedure as before consists in including  $\theta_i$  in the selected group if

$$(y \in R_i) \quad \text{or if} \quad (y_i > C \quad \text{and} \quad y \in R'_i).$$

For this procedure it was shown previously that if  $\theta_i = \Delta$ ,  $0 \leq \theta_j < \Delta$  for  $j \neq i$ , the expected proportion of true positives has been increased by the modification. Suppose now that two of the  $\theta$ 's are equal to  $\Delta$ , say  $\theta_1 = \theta_2 = \Delta$ ,  $0 \leq \theta_j < \Delta$  for  $j > 2$ . Then twice the expected proportion of true positives equals

$$P\{\text{selecting } \theta_1\} + P\{\text{selecting } \theta_2\}.$$

Since for the points  $(\Delta, \Delta, \theta_3, \dots, \theta_a)$  with  $0 \leq \theta_j < \Delta$  for  $j > 2$  we have both  $P(R_1) > P(R'_1)$  and  $P(R_2) > P(R'_2)$ , it is seen that the expected proportion is increased also in this case, and in the same way that it is increased at all points at which it takes on its minimum under (28). The remainder of the argument requires no change.

In conclusion we mention, without going into details, that even without the restriction (30) the procedure (28) is not the solution of the problem of minimizing the maximum expected number of false negatives subject to

$$\sup [\text{Expected number of false positives}] \leq \gamma'.$$

This follows more simply but by the same method as before from the fact that the expected number of false negatives under (28) takes on its maximum at the single point  $\theta_1 = \dots = \theta_a = \Delta$ .

**9. Decision theoretic approach.** Although the formulations of Section 3 appear to the author to be more useful for most applications, the problems can also be treated from a purely decision theoretic point of view, with general loss functions replacing the consideration of true and false positives. In the present section such a treatment of the problems of Sections 5 and 6 will be sketched very briefly.

Suppose that  $X_{i1}, \dots, X_{ia_i}$  ( $i = 1, \dots, a$ ) are independent samples, that the distribution of the  $i$ th sample depends only on the parameter  $\theta_i$  and that we wish to select the populations with high  $\theta$ -values. Let the loss resulting from the selection or nonselection of the  $i$ th population depend only on  $\theta_i$  and be denoted by  $L_i(\theta_i)$  and  $L'_i(\theta_i)$  respectively. Finally, let the over-all loss be the sum of the individual losses.

Consider now the  $i$ th component problem, a two-decision problem for the parameter  $\theta_i$  with losses  $L_i$  and  $L'_i$ . Suppose that the minimax solution  $\psi_i$  for this problem is a Bayes solution with respect to a least favorable *a priori* distri-

bution  $\lambda_i$  of  $\theta_i$ . Then  $\psi_i$  is also the Bayes solution for this same two-decision problem on the basis of all  $\sum n_i$  observations with respect to the *a priori* distribution  $\lambda_1(\theta_1) \times \lambda_2(\theta_2) \times \cdots \times \lambda_a(\theta_a)$  for the combined parameter  $\theta = (\theta_1, \dots, \theta_a)$ . This is an immediate consequence of the fact that the *a posteriori* (marginal) distribution of  $\theta_i$  given all the  $\sum n_i$   $x$ 's depends only on  $x_{i1}, \dots, x_{in_i}$ . It then follows from result (ii) (on p. 15) of [30] that the selection procedure  $(\psi_1^0, \dots, \psi_a^0)$  is a minimax solution of the over-all problem.

Conditions under which the minimax solutions for the component problems are of the form (24) are given for example in [34] and [35]. The minimax property of procedure (24) in these cases is a slight generalization of a result of Robbins [23] and Hannan and Robbins [36], which was established there by quite different methods. It is suggested by these papers (see also Johns [37]) that for the problem under consideration there exist asymptotic subminimax procedures so that for large  $a$ , certain improvements over the above minimax procedure may be possible.

**10. Comparison of normal means with a control.** In the remaining two sections we shall be concerned with problems in which the quality of the standard is not assumed known but where instead a control group  $X_{0j}$  ( $j = 1, \dots, m$ ) is observed in addition to the observations  $X_{ij}$  ( $j = 1, \dots, n_i$ ;  $i = 1, \dots, a$ ) on the  $a$  treatments.

We consider first the case that the  $X_{ij}$  are independently distributed with normal distributions  $N(\xi_i, \sigma_i^2)$  for  $i = 1, \dots, a$  and  $N(\xi_0, \sigma_0^2)$  for  $i = 0$ , and assume to begin with that  $\sigma_0^2, \sigma_1^2$  are known and that  $n_i = n$  for  $i = 1, \dots, a$ . The averages  $\bar{X}_0, \bar{X}_1, \dots, \bar{X}_a$  are then sufficient statistics, independently distributed with normal distributions  $N(\xi_0, \tau_0^2)$  for  $\bar{X}_0$  and  $N(\xi_i, \tau_i^2)$  for  $\bar{X}_i$  ( $i = 1, \dots, a$ ) where

$$\tau_0^2 = \sigma_0^2/m \quad \text{and} \quad \tau_i^2 = \sigma_i^2/n.$$

As in Section 5, a slight generalization of the Hunt-Stein theorem permits a reduction of the data. We may restrict attention to the variables  $Y_i = \bar{X}_i - \bar{X}_0$  since by this theorem there exists a minimax solution which is invariant under a common translation of all variables and since  $Y_1, \dots, Y_a$  constitute a maximal set of invariants with respect to these transformations.

Putting  $\theta_i = \xi_i - \xi_0$ , the conditional joint density of the  $Y$ 's given  $\bar{X}_0 = x_0$  is (up to a constant factor)

$$\exp \left\{ -\frac{1}{2\tau_1^2} \sum [(y_j - \theta_j) + (x_0 - \xi_0)]^2 \right\}.$$

The joint density of the  $Y$ 's is therefore

$$C \int \exp \left\{ -\frac{1}{2\tau_1^2} \sum [(y_j - \theta_j) + y_0]^2 - \frac{1}{2\tau_0^2} y_0^2 \right\} dy_0,$$

which after some simplification becomes

$$(36) \quad p(y) = C \exp \left\{ -\frac{1}{2\tau_1^2} \left[ \sum (y_j - \theta_j)^2 - \frac{a^2 \tau_0^2}{a\tau_0^2 + \tau_1^2} (\bar{y} - \bar{\theta})^2 \right] \right\}.$$

We shall now apply the theorem of Section 4 with  $\theta_0 = 0$  and  $\theta' = \rho\Delta$ . Then  $p_i(y)/p_0(y)$  is the ratio of the two densities (36) for  $\theta_i = \Delta$ ,  $\theta_j = \rho\Delta$  ( $j \neq i$ ) and  $\theta_1 = \dots = \theta_a = 0$ . The quadratic terms in the exponent cancel and the linear terms are, up to a factor  $\Delta/\tau_1^2$ ,

$$\begin{aligned} \rho \sum_{j \neq i} y_j + y_i - \frac{a^2 \tau_0^2}{a\tau_0^2 + \tau_1^2} \frac{1 + (a-1)\rho}{a} \bar{y} \\ = (1-\rho)y_i + a\bar{y} \left\{ \rho - \frac{\tau_0^2}{a\tau_0^2 + \tau_1^2} [1 + (a-1)\rho] \right\}. \end{aligned}$$

For

$$(37) \quad \rho = \frac{\tau_0^2}{\tau_0^2 + \tau_1^2}$$

the coefficient of  $a\bar{y}$  vanishes, so that

$$p_i(y)/p_0(y) = C \exp \left\{ \frac{\Delta(1-\rho)}{\tau_1^2} y_i \right\}$$

is an increasing function of  $y_i$ . For this value of  $\rho$ , the conditions of the theorem of Section 4 are satisfied and the minimax procedure (15) thus reduces to

$$(38) \quad \psi_i = 1 \quad \text{when} \quad y_i = \bar{X}_i - \bar{X}_0 > C,$$

where  $C$  is defined by

$$(39) \quad P_\Delta(\bar{X}_i - \bar{X}_0 > C) = 1 - \Phi \left( \frac{C}{\sqrt{\frac{\sigma_0^2}{m} + \frac{\sigma_1^2}{n}}} - \Delta \right) = \gamma.$$

This solution is easily extended to the case of unequal sample sizes and unequal variances. If the variance of  $\bar{X}_j$  is  $\tau_j^2$  we find for the joint distribution of the  $Y$ 's,

$$p(y) = C \exp \left\{ -\frac{1}{2} \sum_{j=1}^a \frac{1}{\tau_j^2} (y_j - \theta_j)^2 + \frac{1}{2} \frac{[\sum (y_j - \theta_j)/\tau_j^2]^2}{\frac{1}{\tau_0^2} + \sum \frac{1}{\tau_j^2}} \right\}.$$

We now apply the theorem given in Section 6, this time with  $\theta' = \theta'_i = \rho_i\Delta$ , so that  $p_i(y)$  is the density of the  $Y$ 's when

$$\theta_i = \Delta, \quad \theta_j = \rho_i\Delta \quad \text{for } j \neq i.$$

Then the quadratic terms in the exponent again cancel in the ratio  $p_i(y)/p_0(y)$ , and the linear term is

$$\sum \theta_j y_j / \tau_j^2 - \left[ \frac{1}{\tau_0^2} + \sum \frac{1}{\tau_k^2} \right]^{-1} \sum \frac{y_j}{\tau_j^2} \sum \frac{\theta_j}{\tau_j^2}$$

all sums extending from 1 to  $a$ . For  $j \neq i$ , the coefficient of  $y_j$  is (up to a factor  $\Delta[(1/\tau_0^2) + \sum (1/\tau_k^2)]^{-1}$ )

$$\frac{\rho_i}{\tau_j^2} \left[ \frac{1}{\tau_0^2} + \sum \frac{1}{\tau_k^2} \right] - \frac{1}{\tau_j^2} \left[ \frac{1}{\tau_i^2} + \rho_i \sum_{k \neq i} \frac{1}{\tau_k^2} \right].$$

This will be zero if  $\rho_i = \tau_0^2/(\tau_0^2 + \tau_i^2)$  and the coefficient of  $y_j$  then becomes

$$\frac{1}{\tau_i^2} \left[ \frac{1}{\tau_0^2} + \sum_{j \neq i} \frac{1}{\tau_j^2} (1 - \rho_i) \right].$$

Since this is positive,  $p_i(y)/p_0(y)$  is then an increasing function of  $Y_i = \bar{X}_i - \bar{X}_0$  and the minimax procedure is therefore given by (24) where  $C_i$  is given by (39) with  $n_i$  in place of  $n$ , or by (25).

If the variables  $X_{ij}$  ( $j = 1, \dots, n_i$ ;  $i = 0, \dots, a$ ) all have common but unknown variances, it follows as in Section 7 that the procedure given by

$$(40) \quad \psi_i = 1 \quad \text{when} \quad \frac{(\bar{X}_i - \bar{X}_0) / \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}}{\sqrt{\sum_{j=0}^a \sum_{k=1}^{n_j} (X_{jk} - \bar{X}_j)^2}} > C$$

is approximately minimax, where  $n_0$  replaces the earlier  $m$ .

**11. Comparison of normal variances with a control.** Let  $X_{ij}$  ( $j = 1, \dots, n_i$ ;  $i = 0, \dots, a$ ) be independently distributed with normal distributions  $N(\xi_i, \sigma_i^2)$  ( $\xi_i, \sigma_i$  unknown), and consider the problem of selecting the populations for which  $\sigma_i^2/\sigma_0^2 \leq \delta$ . Application of the generalized Hunt-Stein theorem proves the existence of a minimax procedure depending only on the statistics

$$S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2.$$

We may therefore restrict attention to  $S_0^2, \dots, S_a^2$  where the distribution of  $S_i^2/\sigma_i^2$  is  $\chi_{f_i}^2$  with  $f_i = n_i - 1$ , and by another application of the same theorem to the variables  $V_i = S_i^2/S_0^2$  ( $i = 1, \dots, a$ ). The joint densities of the  $V$ 's, is up to a constant factor

$$\left[ \left( 1 + \sum_{j=1}^a \frac{v_j}{\sigma_j^2/\sigma_0^2} \right)^{(f_0+f_1+\dots+f_{a-1})/2} \right]^{-1} \prod_{i=1}^a \frac{v_i^{(f_i-2)/2}}{(2\sigma_i^2/\sigma_0^2)^{f_i/2} \Gamma(f_i/2)}.$$

Let us now apply the theorem of Section 6 with  $\theta_i = \sigma_0^2/\sigma_i^2$ , with  $\theta_0 = 1$  (so that the conditions  $\theta_1 = \dots = \theta_a = \theta_0$  are equivalent to  $\sigma_1^2 = \dots = \sigma_a^2 = \sigma_0^2$ ); with  $1 + \Delta = 1/\delta$  (so that the conditions  $\theta_i \geq \theta_0 + \Delta = 1 + \Delta$  is equivalent to  $\sigma_i^2/\sigma_0^2 \leq \delta$ ); and  $\theta' = \rho_i(1 + \Delta)$ , where  $\rho_i < 1$  is to be determined later. With these values, the probability ratio  $p_i/p_0$  is an increasing function of  $(1 + \sum v_j)/(1 + \sum \theta_j v_j)$ .

Let us therefore consider the region

$$1 + \sum v_j \geq k_i(1 + \sum \theta_j v_j)$$

which is equivalent to

$$k_i(1 - \rho_i)(1 + \Delta)v_i \leq (1 - k_i) + [1 - \rho_i k_i(1 + \Delta)] \sum_{j=1}^a v_j.$$

If we put  $\rho_i = 1/k_i(1 + \Delta)$ , this reduces to

$$v_i \leq \frac{1 - k_i}{k_i(1 - \rho_i)(1 + \Delta)}$$

or equivalently to  $v_i \leq C_i$  with

$$C_i = \frac{1 - k_i}{k_i(1 + \Delta) - 1}.$$

As  $k_i$  goes from  $1/(1 + \Delta)$  to 1,  $C_i$  goes from  $\infty$  to 0, and for these values of  $k_i$ , it is seen that  $\rho_i < 1$ . The theorem of Section 6 therefore shows that the procedure  $\psi_i = 1$  if  $v_i \leq C_i$  has the desired minimax property.

As a last problem, consider a set of Poisson populations with Poisson parameter  $\lambda_0, \lambda_1, \dots, \lambda_a$ . The problem of selecting the populations for which  $\lambda_i/\lambda_0 \geq 1 + \Delta$  is not meaningful in the formulation given here since the parameter pairs  $(\lambda_0, \lambda_i = \lambda_0)$  and  $(\lambda_0, \lambda_i = (1 + \Delta)\lambda_0)$  become indistinguishable as  $\lambda_0 \rightarrow \infty$ . The problem could be treated with the minimax principle replaced by a suitable unbiasedness principle. Alternatively, if one is concerned with Poisson processes, one may instead of observing the number of occurrences in fixed intervals, take as observations the times required to get a specified number of occurrences. These times then follow gamma distributions, and the solution of the present section is directly applicable.

#### REFERENCES

- [1] Z. W. BIRNBAUM, "Effect of linear truncation on a multinormal population," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 272-279.
- [2] Z. W. BIRNBAUM AND D. G. CHAPMAN, "On optimum selections from multinormal populations," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 443-447.
- [3] T. W. ANDERSON, "Classification by multivariate analysis," *Psychometrika*, Vol. 16 (1951), pp. 31-50.
- [4] W. G. COCHRAN, "Improvement by means of selection," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 449-470.
- [5] D. J. FINNEY, "The consequences of selection for a variate subject to errors of measurement," *Rev. Inst. Internat. Stat.*, Vol. 24 (1956), pp. 1-10.
- [6] D. J. FINNEY, "Statistical problems in plant selection," *Bull. Internat. Stat. Inst.*, Vol. 36 (1958), pp. 242-268.
- [7] O. L. DAVIES, "The design of screening tests in the pharmaceutical industry," *Bull. Internat. Stat. Inst.*, Vol. 36 (1958), pp. 226-241.
- [8] R. N. CURNOW, "The consequences of errors of measurement for selection from certain non-normal distributions," *Bull. Internat. Stat. Inst.*, Vol. 37 (1960), pp. 291-308.



- [9] C. W. DUNNETT, "On selecting the largest of  $k$  normal population means," *J. Roy. Stat. Soc., Ser. B*, Vol. 22 (1960), pp. 1-40.
- [10] F. J. ANSCOMBE, "Rejection of outliers," *Technometrics*, Vol. 2 (1960), pp. 123-147.
- [11] FREDERICK MOSTELLER, "A  $k$ -sample slippage test for an extreme population," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 58-65.
- [12] EDWARD PAULSON, "A multiple decision procedure for certain problems in analysis of variance," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 95-98.
- [13] RAGHU RAJ BAHADUR, "On a problem in the theory of  $k$  populations," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 362-375.
- [14] RAGHU RAJ BAHADUR AND LEO A. GOODMAN, "Impartial decision rules and sufficient statistics," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 553-562.
- [15] EDWARD PAULSON, "An optimum solution to the  $k$ -sample slippage problem for the normal distribution," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 610-616.
- [16] R. DOORNBOOS AND H. J. PRINS, "On slippage tests," *Indagationes Mathematicae*, Vol. 20 (1958), pp. 38-55 and pp. 438-447.
- [17] AKIO KUDÔ, "On the invariant multiple decision procedures," *Bull. Math. Stat., Res. Assn. Stat. Sci.*, Vol. 6 (1956), pp. 57-68.
- [18] J. PFANZAGL, "Ein kombiniertes Test und Klassifikations-Problem," *Metrika*, Vol. 2 (1959), pp. 11-45.
- [19] K. V. RAMACHANDRAN AND C. G. KHATRI, "On a decision procedure based on the Tukey statistic," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 802-806.
- [20] DONALD R. TRAU, "An optimum slippage test for the variances of  $k$  normal distributions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 669-674.
- [21] SAMUEL KARLIN AND DONALD TRUAX, "Slippage problems," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 296-324.
- [22] S. S. GUPTA, "On a decision rule for a problem in ranking measures," unpublished Ph.D. Thesis, University of North Carolina.
- [23] HERBERT ROBBINS, "Asymptotically subminimax solutions of compound statistical decision problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 131-148.
- [24] K. C. SEAL, "On a class of decision procedures for ranking means of normal populations," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 387-398.
- [25] K. C. SEAL, "An optimum decision rule for ranking means of normal populations," *Calcutta Stat. Assn. Bull.*, Vol. 7 (1957), pp. 131-150.
- [26] K. C. SEAL, "On ranking parameters of scale in type III populations," *J. Amer. Stat. Assn.*, Vol. 53 (1958), pp. 164-175.
- [27] E. S. PEARSON AND C. CHANDRASEKHAR, "The efficiency of statistical tools and a criterion for the rejection of outlying observations," *Biometrika*, Vol. 28 (1936), pp. 308-320.
- [28] CHARLES M. STEIN, "On the selection of the largest of a set of means," (Abstract), *Ann. Math. Stat.*, Vol. 19 (1948), p. 429.
- [29] ROBERT E. BECHOFER, "A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs," *Biometrics*, Vol. 14 (1958), pp. 408-429.
- [30] E. L. LEHMANN, "A theory of some multiple decision problems, I," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1-25.
- [31] COLIN R. BLYTH, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 22-42.
- [32] E. L. LEHMANN, *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.
- [33] CHARLES W. DUNNETT, "A multiple comparison procedure for comparing several treatments with a control," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1096-1121.

- [34] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.
- [35] SAMUEL KARLIN AND HERMAN RUBIN, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272-299.
- [36] JAMES F. HANNAN AND HERBERT ROBBINS, "Asymptotic solutions of the compound decision problem for two completely specified distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 37-51.
- [37] M. V. JOHNS, JR., "Non-parametric empirical Bayes Procedures," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 649-669.

## BAYES RULES FOR A COMMON MULTIPLE COMPARISONS PROBLEM AND RELATED STUDENT-*t* PROBLEMS<sup>1</sup>

BY DAVID B. DUNCAN

*University of North Carolina<sup>2</sup>*

**0. Summary.** The paper is mainly concerned with the following multiple comparisons problem in the analysis of variance setting. In a balanced experiment  $n$  treatments are to be compared. Each of the  $\frac{1}{2}n(n-1)$  pairwise comparisons is to be made, adjudging each difference as "positive", "negative", or "not significant"; overall decisions involving intransitivities are barred. The loss for each difference is proportional to the error; if a difference is asserted incorrectly the loss has proportionality constant  $c_1$ , if "not-significant" is the incorrect conclusion the proportionality constant is  $c_0$ ; where  $c_1 = k_1 + k_0$ ,  $c_0 = k_0$  and  $k_1 > k_0 > 0$ . Total loss for the experiment is taken as the sum of the  $\frac{1}{2}n(n-1)$  component losses. The Bayes rule for any prior distribution is shown as a result to consist in the simultaneous application of Bayes rules to the  $\frac{1}{2}n(n-1)$  component problems. Each of these in turn is shown similarly to consist in the simultaneous application of Bayes rules to two subcomponent problems. The subcomponent Bayes rule for a normal prior density of treatment means is explicitly derived. The dependencies of the solution on the variance of the prior density, the degrees of freedom and the loss ratio  $k_1/k_0$  are discussed. A principal finding is that the Bayes solution for the multiple comparisons problem corresponds to a tolerated error probability "of the first kind" for each single difference, that is independent of the number of treatments being compared.

**1. Introduction.** Many procedures have been proposed for the multiple comparisons problem herein considered. These include, for example, a "least-significant-difference" rule due to Fisher [5], an "honest-significant-difference" rule due to Tukey [17], [18] and multiple range testing procedures due to Newman [10], and the author [3]. Some of these have also been described in recent texts such as Federer [4], Li [9], Snedecor [13], Scheffé [12] and Steel and Torrie [14]. With much help from the recent more general work of Lehmann [7] it has now been possible to solve a Jeffrey's-like Bayes formulation of the problem. This is more complete than any of the previous formulations and leads to a simple solution with properties that are better defined and that appear to be appropriate to an appreciable class of practical situations. In the process, similar Jeffrey's-like Bayes formulations and their solutions are presented and obtained for two com-

Received November 21, 1960; revised June 16, 1961.

<sup>1</sup> Research sponsored jointly by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command under Contracts No. AF 49(638)-929 and No. AF 49(638)-261, by the Office of Naval Research under contract No. Nonr 355(06), and by the U. S. Public Health Service. Reproduction in whole or in part is permitted for any purpose of the United States Government.

<sup>2</sup> Now at the Johns Hopkins University.

mon types of Student- $t$  problems. These are developed first as problems of separate interest in Sections 2 and 3. The main problem is then fully developed in Section 4. A discussion of the more applied aspects, together with further illustrations is planned for a paper to be submitted to *Biometrics*.

**2. A two-decision Student- $t$  problem.** Given a random observation  $t$  from a non-central  $t$  distribution with non-centrality parameter  $\tau$  and  $v$  degrees of freedom, a common problem is that of choosing between the two decisions

$$(2.1) \quad d_0: \text{decide } \tau \leq \Delta \quad \text{and} \quad d_1: \text{decide } \tau > \Delta,$$

where  $\Delta$  is some unspecified positive boundary value. In the language of the experimenter,  $d_0$  is the decision that  $\tau$  is not significantly greater than zero and  $d_1$  the decision that it is. In the theory of hypothesis testing the same problem is often more loosely regarded as that of testing  $H_0: \tau \leq 0$  with the alternative  $H_1: \tau > 0$ , the decisions thus being

$$(2.2) \quad d_0: \text{decide } \tau \leq 0 \quad \text{and} \quad d_1: \text{decide } \tau > 0$$

Strictly speaking however the null decision does not deny the possibility of positive though relatively small values for  $\tau$  and some such formulation as (2.1) is more precise. The change is relatively trivial in this problem by itself. It is essential however to our subsequent developments as is brought out shortly after (4.12). (See Lehmann [7] also for a similar change).

Our first result is a Bayes rule  $\phi(t)$  for this problem with respect to a simple linear loss function

$$(2.3) \quad \begin{aligned} L_0(\tau) &= L(\tau, d_0) = \begin{cases} 0, & \tau \leq 0, \\ k_0 \tau, & \tau > 0, \end{cases} \\ L_1(\tau) &= L(\tau, d_1) = \begin{cases} k_1 |\tau|, & \tau \leq 0, \\ 0, & \tau > 0, \end{cases} \end{aligned}$$

where  $k_0$  and  $k_1$  are positive constants such that  $k_1 > k_0$ , and with respect to a normal prior density for  $\tau$ ,

$$(2.4) \quad \xi(\tau) = (2\pi\gamma^2)^{-1/2} e^{-\tau^2/2\gamma^2}, \quad -\infty < \tau < \infty,$$

with mean zero and variance  $\gamma^2$ . The rule is of the common form

$$(2.5) \quad \phi_*(t) = \begin{cases} 0, & t < t_*, \\ 1, & t > t_*, \end{cases}$$

where  $\phi(t) = 0$  or 1 is the usual indicator function for making the decision  $d_0$  and  $d_1$  respectively, and  $t_* = t_*(k, v, \gamma^2)$  is a significant or critical  $t$  ratio for which a set of values are given in Table 1. The arguments determining the significant value  $t_*$ , are the ratio  $k = k_1/k_0$  from the loss function, the degrees of freedom  $v$  for  $t$  and the variance  $\gamma^2$  of the prior density for  $\tau$ .

The rule  $\phi_*(t)$  may be derived as follows. For the average risk of any rule

TABLE 1  
Minimum-Average-Risk Significant  $t$  Values ( $t_*$  Values)

$\gamma^2$	$v$	0.0	.5	1.0	Log $k$		2.5	3.0	3.5
					1.5	2.0			
$\infty$	1	0.0	.375	.807	1.353	2.102	3.160	4.685	6.854
	2	0.0	.413	.860	1.379	2.012	2.814	3.851	5.208
	4	0.0	.434	.884	1.367	1.900	2.502	3.197	4.010
	6	0.0	.443	.891	1.356	1.848	2.374	2.948	3.580
	14	0.0	.451	.898	1.340	1.779	2.217	2.654	3.099
	$\infty$	0.0	.457	.902	1.326	1.721	2.091	2.436	2.759
3	1	0.0	.444	1.053	2.503	$\infty$	$\infty$	$\infty$	$\infty$
	2	0.0	.484	1.060	1.926	4.077	$\infty$	$\infty$	$\infty$
	4	0.0	.506	1.056	1.718	2.623	4.178	9.595	$\infty$
	6	0.0	.515	1.053	1.653	2.370	3.308	4.732	7.706
	14	0.0	.522	1.047	1.582	2.136	2.724	3.360	4.074
	$\infty$	0.0	.528	1.041	1.531	1.987	2.414	2.813	3.186
1	1	0.0	.572	1.930	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$
	2	0.0	.610	1.532	8.741	$\infty$	$\infty$	$\infty$	$\infty$
	4	0.0	.629	1.395	2.648	8.592	$\infty$	$\infty$	$\infty$
	6	0.0	.637	1.353	2.303	3.980	13.625	$\infty$	$\infty$
	14	0.0	.642	1.308	2.030	2.859	3.891	5.326	7.818
	$\infty$	0.0	.646	1.275	1.875	2.433	2.957	3.445	3.902
.5	1	0.0	.767	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$
	2	0.0	.785	2.292	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$
	4	0.0	.791	1.963	9.243	$\infty$	$\infty$	$\infty$	$\infty$
	6	0.0	.794	1.800	3.777	$\infty$	$\infty$	$\infty$	$\infty$
	14	0.0	.792	1.653	2.693	4.162	6.670	$\infty$	$\infty$
	$\infty$	0.0	.792	1.562	2.296	2.980	3.622	4.219	4.779

$\phi(t)$ , we have

$$\begin{aligned}
 A(\xi, \phi) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [L_0(\tau)(1 - \phi(t)) + L_1(\tau)\phi(t)]f(t|\tau) dt \xi(\tau) d\tau \\
 (2.6) \quad &= \text{constant} + \int_{-\infty}^{\infty} \phi(t)h_1(t) dt,
 \end{aligned}$$

where  $f(t|\tau)$  is the non-central  $t$  density function, (2.12) below, and

$$(2.7) \quad h_1(t) = \int_{-\infty}^{\infty} [L_1(\tau) - L_0(\tau)]f(t|\tau)\xi(\tau) d\tau.$$

The minimum average risk rule may thus be written

$$(2.8) \quad \phi_*(t) = \begin{cases} 0, & h_1(t) > 0, \\ 1, & h_1(t) < 0. \end{cases}$$

From the inequality  $h_1(t) < 0$  we get

$$(2.9) \quad \int_{-\infty}^0 k_1 |f(t|\tau)| \xi(\tau) d\tau < \int_0^{\infty} k_0 \tau f(t|\tau) \xi(\tau) d\tau,$$

and hence

$$(2.10) \quad h_2(t) = \frac{\int_0^{\infty} h_3(\tau, t) d\tau}{-\int_{-\infty}^0 h_3(\tau, t) d\tau} > \frac{k_1}{k_0} = k,$$

where

$$(2.11) \quad h_3(\tau, t) = \tau f(t|\tau) \xi(\tau).$$

Now, introducing the non-central  $t$  density in the form

$$(2.12) \quad f(t|\tau) = \int_0^{\infty} \frac{e^{-\frac{1}{2}(ut-\tau)^2}}{(2\pi)^{\frac{1}{2}}} u \psi(u|v) du, \quad -\infty < t < \infty,$$

where  $\psi(u|v)$  is the  $\chi^2$ -related density function of  $u = \chi(v)/v$ , that is

$$(2.13) \quad \begin{aligned} \psi(u|v) &= (v^{\frac{1}{2}} u^{v-1} e^{-\frac{1}{2}vu^2}) / [(\frac{1}{2}v-1)! 2^{\frac{1}{2}v-1}], & u > 0, \\ &= 0, & \text{otherwise,} \end{aligned}$$

and discarding constants in  $h_3(\tau, t)$  which will cancel and not affect the value of  $h_2(t)$ , we have

$$(2.14) \quad h_3(\tau, t) \propto \tau \int_0^{\infty} \exp\{\frac{1}{2}[(ut-\tau)^2 + \tau^2/\gamma^2]\} u \psi(u|v) du.$$

Putting  $\beta^2 = 1 + 1/\gamma^2$  this becomes

$$(2.15) \quad \begin{aligned} h_3(\tau, t) &\propto \tau e^{-\frac{1}{2}\tau^2\beta^2} \int_0^{\infty} \exp\{ut\tau - \frac{1}{2}u^2(v+t^2)\} u^v du \\ &= \tau e^{-\frac{1}{2}\tau^2\beta^2} \sum_{i=0}^{\infty} \frac{(t\tau)^i}{i!} \int_0^{\infty} u^{v+i} \exp\{-\frac{1}{2}u^2(v+t^2)\} du \\ &= \tau e^{-\frac{1}{2}\tau^2\beta^2} \sum_{i=0}^{\infty} \frac{(t\tau)^i}{i!} \left(\frac{2}{v+t^2}\right)^{\frac{1}{2}i} \left(\frac{v+i-1}{2}\right)!. \end{aligned}$$

Next, putting  $h_4(t) = \int_0^{\infty} h(\tau, t) d\tau$  and  $y = t/\beta(v+t^2)^{\frac{1}{2}}$  and integrating term by term, we get

$$(2.16) \quad \begin{aligned} h_4(t) &\propto \sum_{i=0}^{\infty} \frac{y^i}{i!} 2^{i/2} \left(\frac{v+i-1}{2}\right)! \int_0^{\infty} (\tau\beta)^{i+1} e^{-\frac{1}{2}(\tau\beta)^2} d(\tau\beta) \\ &\propto \sum_{i=0}^{\infty} \frac{y^i}{i!} 2^i \left(\frac{v+i-1}{2}\right)! \left(\frac{i}{2}\right)! \\ &= \sum_{i=0}^{\infty} \frac{y^i}{i!} 2^i \left(\frac{i+1}{2}\right)! \left(\frac{i}{2}\right)! \left(\frac{v+i-1}{2}\right)! / \left(\frac{i+1}{2}\right)! \end{aligned}$$



But  $2^i[(i+1)/2]!(i/2)! = (i+1)! \frac{1}{2}(\pi)^{\frac{1}{2}}$ , from which

$$(2.17) \quad \begin{aligned} h_4(t) &= \frac{d}{dy} \left[ y \sum_{i=0}^{\infty} y^i \left( \frac{v+i-1}{2} \right)! / \left( \frac{i+1}{2} \right)! \right] \\ &= \frac{d}{dy} [y h_5(y) + y h_6(y)], \end{aligned}$$

where  $h_5(y)$  is the sum of even terms,

$$h_5(y) = \sum_{i=0}^{\infty} y^{2i} (p+i-\frac{1}{2})! / (i+\frac{1}{2})!,$$

with  $p = v/2$ , and  $h_6(y)$  is the sum of odd terms,

$$h_6(y) = \sum_{i=0}^{\infty} y^{2i+1} (p+i)! / (i+1)!.$$

Working first with  $h_5(y)$ , it may be written as

$$(2.18) \quad h_5(y) = (p - \frac{1}{2})! F(p + \frac{1}{2}, 1; \frac{3}{2}; y^2) / (\frac{1}{2})!,$$

where  $F(a, b; c; x)$  denotes the hypergeometric function

$$1 + \frac{abx}{c} + \frac{a(a+1)b(b+1)}{c(c+1)} \frac{x^2}{2} + \dots$$

Applying Euler's transformation

$$F(a, b; c; x) = (1-x)^{c-a-b} F(c-a, c-b; c; x)$$

and reducing, we get

$$(2.19) \quad h_5(y) = \frac{1}{y} (1-y^2)^{-p} \int_0^y (1-u^2)^{p-1} du (p - \frac{1}{2})! / (\frac{1}{2})!.$$

Hence

$$(2.20) \quad \frac{d}{dy} [y h_5(y)] = (1-y^2)^{-(p+1)} \left[ (1-y^2)^p + 2py \int_0^y (1-u^2)^{p-1} du \right] \cdot (p - \frac{1}{2})! / (\frac{1}{2})!.$$

Next,  $h_6(y)$  sums to  $(p-1)![(1-y^2)^{-p} - 1]/y$  so that

$$(2.21) \quad (d/dy)[y h_6(y)] = (1-y^2)^{-(p+1)} 2p! y.$$

Treating the denominator of  $h_2(t)$  in the same way and combining results we get

$$(2.22) \quad h_2(t) = h_4(t)/h_4(-t) = g(y)/g(-y),$$

where

$$(2.23) \quad g(y) = (1-y^2)^{1/2} + vy \int_0^y (1-u^2)^{1/2(v-2)} du + \frac{2(\frac{1}{2})!(\frac{v}{2})! y}{[\frac{1}{2}(v-2)]!}.$$

From (2.8), (2.10) and (2.22) we have

$$(2.24) \quad \phi_*(t) = \begin{cases} 0, & g(y)/g(-y) < k, \\ 1, & g(y)/g(-y) > k. \end{cases}$$

Hence, since  $g(y)/g(-y)$  is monotone increasing with respect to  $y$ ,

$$(2.25) \quad \phi_*(t) = \begin{cases} 0, & y < y_*, \\ 1, & y > y_*, \end{cases}$$

where  $y_* = y_*(k, v)$  is the solution for  $y$  in  $g(y)/g(-y) = k$ . Finally, putting  $t_* = t_*(k, v, \gamma^2)$  for the solution for  $t$  in  $y_* = t/\beta(v + t^2)^{1/2}$  we have

$$(2.26) \quad \phi_*(t) = \begin{cases} 0, & t < t_*, \\ 1, & t > t_*, \end{cases}$$

as was to be shown. More specific details of the computation of the significant  $t$  ratios in Table 1 are given in Section 6.

EXAMPLE 1. To illustrate suppose the following: A standard treatment is modified in the hopes of producing an increased yield. An experiment is run giving  $r$  yield observations  $x_{11}, \dots, x_{1r}$  and  $x_{21}, \dots, x_{2r}$  for the new and control (standard) treatment respectively. It can be assumed that the respective sets of data are random independent samples from normal populations with means  $\mu_1$  and  $\mu_2$  and with the same, but unknown, variance  $\sigma^2$ . It is required to decide whether the new treatment is significantly superior (in yield) or not significantly superior than the standard; whether to generally recommend it as the superior or to withhold such a general recommendation. Type-1-like errors of recommending a non-superior new treatment (making  $d_1$  when  $\delta \leq 0$ , where  $\delta = \mu_1 - \mu_2$ ) are thought to increase in seriousness in direct proportion to the degree,  $-\delta$ , of inferiority involved. Type-2-like errors of failing to recommend a superior new treatment (making  $d_0$  when  $\delta > 0$ ) are similarly thought to increase in seriousness in direct proportion to the degree,  $\delta$ , of superiority involved. For any absolute difference  $\delta_0 = |\delta|$ , recommendation of an inferior new treatment with  $\delta = -\delta_0$  is considered  $k$  times as serious as the corresponding failure to recommend a superior new treatment with  $\delta = \delta_0$ . In the averaging of risks it is desired to weight risks symmetrically at  $\delta = \pm\delta_0$  for all possible differences  $\delta_0 \geq 0$ , with weights decreasing with respect to  $\delta_0$  as given by a normal density for  $\tau = \delta/(2\sigma^2/r)^{1/2}$  with mean zero and variance  $\gamma^2$ . A minimum-average-risk rule is required which would be invariant with respect to any changes of scale or location in the observations.

Because of sufficiency and invariance considerations the required rule can be restricted to depend on the observations through only the  $t$  ratio

$$(2.27) \quad t = (\bar{x}_1 - \bar{x}_2)/(2s^2/r)^{1/2},$$

where  $\bar{x}_1$  and  $\bar{x}_2$  are the respective sample means and  $s^2$  is the pooled within-sample variance estimate

$$(2.28) \quad s^2 = \sum_{i=1}^2 \sum_{j=1}^r (x_{ij} - \bar{x}_i)^2 / 2(r-1).$$

The required rule is then given by (2.26) where  $t_* = t_*(k, v, \gamma^2)$  with  $v = 2(r - 1)$ .

In practice a similar problem could well arise in which it is desired to use a non-zero mean  $\mu_3$  for the prior density. With infinite error degrees of freedom ( $v = \infty$ ) the significant  $t$  ratio for such an asymmetric problem can be shown to be given by subtracting a correction of  $\mu_3/(2\sigma^2/r)^{1/2}$  from the corresponding value in Table 1. For  $v$  finite the derivation is more difficult. The use of a similar correction of  $\mu_3/(2s^2/r)^{1/2}$  would no doubt suffice however for practical purposes except for very small values of  $v$ . Since the extensions of this problem in the later sections concern only the symmetric case, a more detailed treatment of the asymmetric case will not be taken up here.

From the roles they play the parameters  $k, v$ , and  $\gamma^2$  determining the minimum-average-risk significant  $t$  values may be usefully termed the *loss* or *error seriousness ratio*, the *error degrees of freedom* and the *risk-weighting variance ratio* respectively. Before going on it is of interest to note in Table 1 that a loss ratio of 100 ( $\log k = 2$ ) infinite error degrees of freedom ( $v = \infty$ ), and a risk weighting variance ratio of 3 ( $\gamma^2 = 3$ ) give a  $t_*$  of 1.987 close to that 1.960 of a .025 level test of  $H_0: \tau \leq 0$ .

**3. A related three-decision Student- $t$  problem.** Given a similar observed  $t$  value, a problem related to that of Section 2 is one of choosing between the three common decisions

$$(3.1) \quad d_0: \text{decide } |\tau| \leq \Delta, \quad d_1: \text{decide } \tau > \Delta \quad \text{and} \quad d_2: \text{decide } \tau < -\Delta,$$

where, as before,  $\Delta$  is some unspecified positive boundary value. In the language of the experimenter  $d_0, d_1$  and  $d_2$  are the decisions that  $\tau$  is not significantly different from zero, that  $\tau$  is significantly greater than zero and that  $\tau$  is significantly less than zero, respectively.

Our second result is a Bayes rule for this problem with respect to a similar linear loss function

$$(3.2) \quad \begin{aligned} L_0^{(2)}(\tau) &= L^{(2)}(\tau, d_0) = \begin{cases} 0, & \tau = 0, \\ c_0 |\tau|, & \tau \neq 0, \end{cases} \\ L_1^{(2)}(\tau) &= L^{(2)}(\tau, d_1) = \begin{cases} c_1 |\tau|, & \tau \leq 0, \\ 0, & \tau > 0, \end{cases} \\ L_2^{(2)}(\tau) &= L^{(2)}(\tau, d_2) = \begin{cases} 0, & \tau < 0, \\ c_1 \tau, & \tau \geq 0, \end{cases} \end{aligned}$$

where  $c_0$  and  $c_1$  are positive constants such that  $c_1 - c_0 > c_0$  and with respect to the same normal prior density (2.4) for  $\tau$ . The rule is

$$(3.3) \quad \phi_*^2(t) = (\phi_{0*}^2(t) \phi_{1*}^2(t) \phi_{2*}^2(t)) = \begin{cases} (1 \ 0 \ 0), & |t| < t_*, \\ (0 \ 1 \ 0), & t > t_*, \\ (0 \ 0 \ 1), & t < -t_*, \end{cases}$$

where the significant  $t$  ratio  $t_* = t_*(k, v, \gamma^2)$  is the same as that of the previous

section with the loss ratio now given by  $k = (c_1/c_0) - 1$ , and where  $\phi_i^{(2)}(t) = 0$  or 1 denotes the not making or making of the decision  $d_i$ ,  $i = 0, 1, 2$ .

This result can be obtained as follows: First the three-decision subset system

$$(3.4) \quad \omega_0: |\tau| \leq \Delta, \quad \omega_1: \tau > \Delta, \quad \omega_2: \tau < -\Delta,$$

can be expressed as the restricted product (the full product less empty intersections, see Lehmann [7]) of two component two-decision subset systems like that of the previous problem in Section 2, namely

$$(3.5) \quad \begin{aligned} &\text{Component system for } +\tau: \omega_0^+: \tau \leq \Delta, \quad \omega_1^+: \tau > \Delta, \\ &\text{Component system for } -\tau: \omega_0^-: -\tau \leq \Delta, \quad \omega_1^-: -\tau > \Delta. \end{aligned}$$

Thus

$$(3.6) \quad \omega_0 = \omega_0^+ \cap \omega_0^-, \quad \omega_1 = \omega_1^+ \cap \omega_1^-, \quad \omega_2 = \omega_0^+ \cap \omega_1^-.$$

The intersection  $\omega_1^+ \cap \omega_1^-$  is excluded since it is empty. Put in other words, each of the main decisions is equivalent to two joint component decisions

$$(3.7) \quad d_0 \text{ to } d_0^+ \text{ with } d_0^-, \quad d_1 \text{ to } d_1^+ \text{ with } d_0^- \text{ and } d_2 \text{ to } d_0^+ \text{ with } d_1^-;$$

the joint decision  $d_i^+$  with  $d_1^-$  is excluded since it has mutually incompatible components;  $d_i^+$  is the decision  $\tau \in \omega_i^+$ ;  $\alpha = +, -$ ;  $i = 0, 1$ .

Second, by putting  $k_1 = c_1 - c_0$  and  $k_0 = c_0$  the losses for the main decisions can be expressed as the sums of losses for its component decisions as given by the two-decision loss function (2.3) in the previous section. Demonstrating this

$$(3.8) \quad \begin{aligned} L_0(\tau) + L_0(-\tau) &= \begin{cases} 0 + k_0(-\tau) = c_0|\tau|, & \tau < 0 \\ 0 + 0 = 0, & \tau = 0 \\ k_0\tau + 0 = c_0|\tau|, & \tau > 0 \end{cases} = L_0^{(2)}(\tau), \\ L_1(\tau) + L_0(-\tau) &= \begin{cases} k_1|\tau| + k_0(-\tau) = c_1|\tau|, & \tau < 0 \\ k_1|\tau| + 0 = c_1|\tau|, & \tau = 0 \\ 0 + 0 = 0, & \tau > 0 \end{cases} = L_1^{(2)}(\tau), \\ L_0(\tau) + L_1(-\tau) &= \begin{cases} 0 + 0 = 0, & \tau < 0 \\ 0 + k_1|-\tau| = c_1\tau, & \tau = 0 \\ k_0\tau + k_1|-\tau| = c_1\tau, & \tau > 0 \end{cases} = L_2^{(2)}(\tau). \end{aligned}$$

Next, any rule  $\phi^{(2)}(t)$  for the three-decision problem can also be expressed in terms of two component two-decision rules. For this purpose it is convenient to first re-express the two-decision function  $\phi(t)$  in the two-element vector form

$$(3.9) \quad \phi(t) = (\phi_0(t) \phi_1(t)),$$

where  $\phi_0(t) = 1 - \phi(t)$  and  $\phi_1(t) = \phi(t)$ . In this form, for example, the Bayes rule  $\phi_*(t)$  of the previous section appears as

$$(3.10) \quad \phi_*(t) = \begin{cases} (0 \ 1), & t < t_*, \\ (1 \ 0), & t > t_*. \end{cases}$$

With this vector notation (3.9) for the two-decision function and the components (3.7) of the three decisions in mind we can write

$$\begin{aligned} \phi^{(2)}(t) &= (\phi_0^{(2)}(t) \quad \phi_1^{(2)}(t) \quad \phi_2^{(2)}(t)) \\ (3.11) \quad &= (\phi_0^+(t)\phi_0^-(t) \quad \phi_1^+(t)\phi_0^-(t) \quad \phi_0^+(t)\phi_1^-(t)), \end{aligned}$$

where  $\phi_i^\alpha(t) = 0$  or 1 denotes the not making or making of decision  $d_i^\alpha$ ;  $\alpha = +, -$ ;  $i = 0, 1$ .

Now, working with the average risk of  $\phi^{(2)}(t)$  we have

$$\begin{aligned} A(\xi, \phi^{(2)}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{i=0}^2 L_i^{(2)}(\tau) \phi_i^{(2)}(t) f(t|\tau) dt \xi(\tau) d\tau \\ (3.12) \quad &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{i=0}^1 \sum_{j=0}^1 (L_i(\tau) + L_j(-\tau)) \phi_i^+(t) \phi_j^-(t) f(t|\tau) dt \xi(\tau) dt, \end{aligned}$$

provided that the condition

$$(3.13) \quad \phi_1^+(t)\phi_1^-(t) = 0, \quad -\infty < t < \infty,$$

is satisfied. (Following Lehmann [7] this may be termed a compatibility condition since, if it were not satisfied, the component rules would give incompatible decisions). Assuming pro tempore that this is satisfied, the average risk readily reduces to

$$(3.14) \quad A(\xi, \phi^{(2)}) = A(\xi, \phi^+(t)) + A(\xi, \phi^-(t)),$$

where the component average risks  $A(\xi, \phi^\alpha(t))$ ,  $\alpha = +, -$ , are the same functions of  $t$  and  $-t$  as was  $A(\xi, \phi) = A(\xi, \phi(t))$  of  $t$  in Section 2. To minimize (3.14) it is sufficient to minimize the component average risks separately subject still, of course, to the compatibility condition (3.13). From the result of the previous section, as expressed in (3.10), the component solutions are

$$(3.15) \quad \phi_+^+(t) = \begin{cases} (1, 0), & t < t_* \\ (0, 1), & t > t_* \end{cases} \quad \text{and} \quad \phi_-^-(t) = \begin{cases} (1, 0), & -t < t_* \\ (0, 1), & -t > t_* \end{cases}$$

Since  $\phi_{1*}^+(t) = 1$  only in  $t > t_*$ ,  $\phi_{0*}^+(t) = 1$  only in  $-t > t_* = t < -t_*$ , and since  $t_*$  is positive (from  $k_1 > k_0$ ) we do have  $\phi_{1*}^+(t)\phi_{1*}^-(t) = 0$  for all  $t$ , that is, the solutions are compatible. Thus the required Bayes rule is given by

$$\begin{aligned} \phi^{(2)}(t) &= (\phi_{0*}^+(t) \quad \phi_{0*}^-(t)\phi_{1*}^+(t)\phi_{0*}^-(t) \quad \phi_{0*}^+(t)\phi_{1*}^-(t)) \\ (3.16) \quad &= \begin{cases} (1, 0, 0), & (t < t_*)(t > -t_*) = |t| < t_*, \\ (0, 1, 0), & (t > t_*)(t > -t_*) = t > t_*, \\ (0, 0, 1), & (t < t_*)(t < -t_*) = t < -t_*, \end{cases} \end{aligned}$$

as was to be shown.

EXAMPLE 2. Suppose that two samples of yields have been observed as in Example 1 except that now they are for two new treatments. It is required to decide whether the first can be recommended as the superior (in yield), whether

the second can, or whether to withhold recommendations on both. Errors of making a wrong recommendation or of failing to make an appropriate recommendation are again scorable in direct proportion to the degree of inferiority or superiority involved respectively. For any difference that may exist the error of a wrong recommendation is considered  $c$  times as serious as that of just failing to make the right recommendation. The requirements for weighting of risks and invariance are the same. The required rule is then given by (3.16) where  $t$  is as obtained (2.27) in Example 1,  $t_* = t_*(k, v, \gamma^2)$  with  $k = c - 1$  and  $v = 2(r - 1)$ . The subtraction of one from the loss ratio  $c$  will be trivial and unnecessary in most practical situations with  $c$  not small. The need for it here and not in Example 1, it may be said, comes from the fact that a wrong recommendation now includes implicitly a failure to make an appropriate recommendation as well.

**4. A symmetric multiple comparisons problem.** Given  $N = \frac{1}{2}n(n - 1)$   $t$  statistics of the form

$$(4.1) \quad t_{pq} = (\hat{\mu}_p - \hat{\mu}_q) / s_{\hat{\mu}_p - \hat{\mu}_q}, \quad pq \in N,$$

with non-centrality parameters of the form

$$(4.2) \quad \tau_{pq} = (\mu_p - \mu_q) / \sigma_{\hat{\mu}_p - \hat{\mu}_q}, \quad pq \in N,$$

where  $N$  is used for convenience to denote the set of pairs

$$\{1, 2; 1, 3; \dots; (n - 1), n\}$$

as well as its size, a common multiple comparisons problem is that of choosing between the three decisions

$$(4.3) \quad d_{pq}^0: \tau_{pq} \in \omega_{pq}^0, \quad d_{pq}^1: \tau_{pq} \in \omega_{pq}^1, \quad d_{pq}^2: \tau_{pq} \in \omega_{pq}^2,$$

simultaneously for all  $pq \in N$ , where the subsets are of the previous (3.1) form

$$(4.4) \quad \omega_{pq}^0: |\tau_{pq}| \leq \Delta, \quad \omega_{pq}^1: \tau_{pq} > \Delta, \quad \omega_{pq}^2: \tau_{pq} < -\Delta.$$

The joint density of the  $t_{pq}$ 's is the one that would result, for example, from the common assumptions (a)  $\hat{\mu}_1, \dots, \hat{\mu}_n$  are normal independent variables with means  $\mu_1, \dots, \mu_n$  and the same but unknown variance  $\sigma_{\hat{\mu}}^2$  and (b)  $s_{\hat{\mu}}$  is an estimator of  $\sigma_{\hat{\mu}}$  with  $v$  degrees of freedom such that  $u = s_{\hat{\mu}}/\sigma_{\hat{\mu}}$  has the  $\chi^2$ -related density (2.13) independently of  $\hat{\mu}_1, \dots, \hat{\mu}_n$ , from which  $\sigma_{\hat{\mu}_p - \hat{\mu}_q} = 2^{1/2} s_{\hat{\mu}}$  and  $s_{\hat{\mu}_p - \hat{\mu}_q} = 2^{1/2} s_{\hat{\mu}}$ . If we put  $\mathbf{y}$  for any vector of  $(n - 1)$  orthogonal normalized comparisons among the estimates  $\hat{\mu}_i$ , that is,  $\mathbf{y} = \mathbf{A}\hat{\boldsymbol{\mu}}$  where  $\mathbf{A}\mathbf{A}' = \mathbf{I}_{n-1}$ ,  $\mathbf{A}\mathbf{j} = \mathbf{0}$  ( $\mathbf{j}$  being a vector of ones), and  $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1 \dots \hat{\mu}_n)'$ , then the density of the  $t_{pq}$ 's can be represented conveniently in terms of that of the  $(n - 1)$ -element  $\mathbf{t}$  vector

$$(4.5) \quad \mathbf{t} = \mathbf{y}/s_{\hat{\mu}}$$

depending on the corresponding  $(n - 1)$ -element non-centrality parameter vector

$$(4.6) \quad \boldsymbol{\tau} = \mathbf{n}/\sigma_{\hat{\mu}}$$



where  $\mathbf{n} = \mathbf{A}\mathbf{u}$  and  $\mathbf{u} = (\mu_1 \cdots \mu_n)'$ . This can readily be expressed as

$$(4.7) \quad f_n(\mathbf{t} | \boldsymbol{\tau}, v) = \int_0^\infty \exp \left\{ -\frac{1}{2}(\mathbf{u}\mathbf{t} - \boldsymbol{\tau})'(\mathbf{u}\mathbf{t} - \boldsymbol{\tau}) \right\} \frac{u^{n-1} \psi(u|v) du}{(2\pi)^{\frac{1}{2}(n-1)}} \quad -\infty < \mathbf{t} < \infty$$

where  $\psi(u|v)$  is the  $\chi^2$ -related density (2.13).

Our main result is a Bayes rule for this problem with respect to generalizations of the linear additive loss functions and normal prior density used in the previous sections.

More explicitly, the subsets  $\omega_0, \omega_1, \dots, \omega_{M-1}$  of the multiple comparisons problem are the non-empty intersections

$$(4.8) \quad \omega_i = \bigcap_{p,q \in N} \omega_{pq}^{j_{pq}}, \quad j_{pq} = 0, 1 \text{ or } 2, \quad i = 0, \dots, M-1,$$

of the subsets of all the component three-decision problems involved. The decision system consists of the  $M$  corresponding decisions

$$(4.9) \quad d_i: \boldsymbol{\tau} \in \omega_i, \quad i = 0, \dots, M-1.$$

For example, thinking of the component subset systems (4.4) in the form

$$(4.10) \quad \omega_{pq}^0: (|\mu_p - \mu_q| \leq \Delta'), \quad \omega_{pq}^1: (\mu_q < \mu_p - \Delta'), \quad \omega_{pq}^2: (\mu_p < \mu_q - \Delta'),$$

(where  $\Delta' = \Delta 2^{\frac{1}{2}} \sigma_\mu$ ) and using the corresponding more graphic notation

$$(4.11) \quad (\underline{p}, \underline{q}), (\underline{q}, \underline{p}), (\underline{p}, \underline{q}),$$

in place of  $\omega_{pq}^0, \omega_{pq}^1, \omega_{pq}^2$  respectively, the  $M = 19$  multiple comparisons subsets

	(1, 2)	(2, 1)	(1, 2)
(2, 3)	$\omega_0 = (1, 2, 3)$	$\omega_1 = (\underline{2}, \underline{3}, 1)$	$\omega_2 = (1, \underline{3}, 2)$
(1, 3) (3, 2)	$\omega_3 = (\underline{3}, 1, 2)$	.	$\omega_4 = (\underline{1}, \underline{3}, 2)$
(2, 3)	$\omega_5 = (\underline{2}, 1, 3)$	$\omega_6 = (2, 1, 3)$	.
(2, 3)	$\omega_7 = (3, 2, 1)$	$\omega_8 = (3, 2, 1)$	.
(3, 1) (3, 2)	$\omega_9 = (\underline{3}, \underline{1}, 2)$	$\omega_{10} = (3, 2, 1)$	$\omega_{11} = (3, 1, 2)$
(2, 3)	.	$\omega_{12} = (2, 3, 1)$	.
(2, 3)	$\omega_{13} = (\underline{1}, 2, 3)$	.	$\omega_{14} = (1, 2, 3)$
(1, 3) (3, 2)	.	.	$\omega_{15} = (1, 3, 2)$
(2, 3)	$\omega_{16} = (\underline{1}, 2, 3)$	$\omega_{17} = (2, 1, 3)$	$\omega_{18} = (1, 2, 3)$

in the case  $n = 3$  may be developed as in (4.12). In general, following Duncan [3], the notation  $(i, j, k, \dots)$  may be used to denote subsets in which the corresponding means  $\mu_i, \mu_j, \mu_k, \dots$  are ranked in significantly ascending order from left to right (i.e. with differences  $|\delta| > \Delta'$ ) except that subscripts underscored by a common line denote pairs of means for which the difference is not significant

(i.e.  $|\delta| \leq \Delta'$ ). Thus  $\omega_2 = (1, 3, 2)$  is the subset in which  $\mu_1$  is significantly less than  $\mu_2$  but  $\mu_1$  and  $\mu_2$  each do not differ significantly from  $\mu_3$ . The remaining  $2^3 - 19 = 8$  intersections, e.g.,  $(2, 1)(1, 3)(3, 2)$ , of the component subsets are not included in the multiple comparisons system since they are empty.

Choosing the elements of  $\tau$  as  $\tau_1 = (\mu_1 - \mu_2)/2^{\frac{1}{2}}\sigma_\mu$  for abscissa and  $\tau_2 = (\mu_1 + \mu_2 - 2\mu_3)/6^{\frac{1}{2}}\sigma_\mu$  for ordinate the parameter subsets, in the case  $n = 3$ , may be represented as shown in Figure 2 of [3], where the vertical lines are  $\tau_{12} = \pm\Delta$ , the lines from top left to bottom right are  $\tau_{13} = \pm\Delta$  and those from bottom left to top right are  $\tau_{23} = \pm\Delta$ .

Referring back for a moment, the need for the definition (2.1) instead of (2.2) for the decisions of our initial subcomponent problem can now be seen more clearly in the formation of the subsets (4.12). If (2.2) were used the six subsets  $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$  and  $\omega_{13}$  of the form  $(i, j, k)$  would be eliminated. These however are useful members of the system, hence the need for some such definition as (2.1) to retain them.

The size  $M$  of the multiple comparisons subset system increases rapidly with  $n$  the number of means involved. In the next case  $n = 4$ , for example, the numbers of the various forms of subsets, using the same notation, are

$$(4.13) \quad \begin{aligned} & (1, 2, 3, 4) \cdots 1, (i, j, k, l) \cdots 4, (i, j, k, l) \cdots 4, (i, j, k, l) \\ & \cdots 12, (i, j, k, l) \cdots 12, (i, j, k, l) \cdots 12, (i, j, k, l) \cdots 24, \\ & (i, j, k, l) \cdots 12, (i, j, k, l) \cdots 12, (i, j, k, l) \cdots 12, (i, j, k, l) \\ & \cdots 24, (i, j, k, l) \cdots 6, (i, j, k, l) \cdots 24, (i, j, k, l) \cdots 24, \end{aligned}$$

making  $M = 183$  in all.

The losses are defined as the sum of the losses (3.2) for each of the component decisions involved; that is

$$(4.14) \quad L_i^{(n)}(\tau) = L^{(n)}(\tau, d_i) = \sum_{j \neq q \in N} L_{j \neq q}^{(2)}(\tau_{j \neq q}),$$

$$j_{j \neq q}^i = 0, 1 \text{ or } 2; i = 0, \dots, M-1.$$

Thus, suppose for example in the case  $n = 3$ ,  $\mathbf{y}/2^{\frac{1}{2}}\sigma = (\mu_1\mu_2\mu_3)'/2^{\frac{1}{2}}\sigma$  represents the expected standardized yields of three manurial treatments on a particular agricultural crop. Then the loss  $L_{14}^{(3)}(\tau)$  at  $\mathbf{y}/2^{\frac{1}{2}}\sigma = (10, 12, 8)'$  incurred by the decision  $d_{14}: \tau \in (1, 2, 3)$  is

$$(4.15) \quad \begin{aligned} L_{14}^{(3)}(\tau) &= L_2^{(2)}(\tau_{12}) + L_2^{(2)}(\tau_{13}) + L_0^{(2)}(\tau_{23}) \\ &= L_2^{(2)}(-2) + L_2^{(2)}(2) + L_0^{(2)}(4) \\ &= 0 + 2c_1 + 4c_0 = 2k_1 + 6k_0. \end{aligned}$$

The third contribution  $4c_0 = 4k_0$  enters, it may be said, because the decision  $d_{14}$  has failed to recognize the 4-unit superiority of treatment two over treatment three. The second contribution  $2c_1 = 2k_1 + 2k_0$  enters, because, in similar terms,  $d_{14}$  not only fails to recognize the 2-unit superiority of treatment one over treatment three, incurring a loss of  $2k_0$ , it also commits the more serious error of

ranking treatment three above treatment one, incurring an additional loss of  $2k_1$ . No loss is incurred by the first component decision because  $d_{14}$  ranks treatment two correctly above treatment one.

The prior density for averaging risks over all points in the  $(n-1)$ -dimensional space for  $\tau$  is the simple normal density function

$$(4.16) \quad \xi_n(\tau | \gamma^2) = (2\pi\gamma^2)^{-1(n-1)} e^{-\tau^2/2\gamma^2}, \quad -\infty < \tau < \infty.$$

This is one which would result, for example, from assuming that the means  $\mu_1, \dots, \mu_n$  have independent normal prior distributions with the same mean  $\theta$  and same variance  $\sigma_\mu^2 = \gamma^2 \sigma_\beta^2$ .

From the additive losses assumption it follows as before that the average risk for any decision rule  $\phi^{(n)}(t) = (\phi_0^{(n)}(t) \dots \phi_{M-1}^{(n)}(t))$  may be expressed as the sum of average risks for component three-decision rules

$$\phi^{pq}(t) = (\phi_0^{pq}(t) \phi_1^{pq}(t) \phi_2^{pq}(t))$$

provided again that the component rules are compatible. The steps may be written

$$\begin{aligned} A(\xi_n, \phi^{(n)}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{i=0}^{M-1} L_i^{(n)}(\tau) \phi_i^{(n)}(t) f_n(t | \tau) dt \xi_n(\tau) d\tau \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j_1=0}^2 \dots \sum_{j_{N-1}=0}^2 [L_{j_1}^{(2)}(\tau_{12}) + \dots + L_{j_{N-1}}^{(2)}(\tau_{(n-1)n})] \\ &\quad \cdot \phi_{j_1}^{12}(t) \dots \phi_{j_{N-1}}^{(n-1)n}(t) f_n(t | \tau) dt \xi_n(\tau) d\tau \\ &= \sum_{pq \in N} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j=0}^2 L_j^{(2)}(\tau_{pq}) \phi_j^{pq}(t) f_n(t | \tau) dt \xi_n(\tau) d\tau \\ &= \sum_{pq \in N} A(\xi_n \phi^{pq}(t)). \end{aligned} \quad (4.17)$$

The compatibility condition may be written as

$$(4.18) \quad \prod_{pq \in N} \phi_{j_{pq}}^{pq}(t) = 0, \quad j_{pq} = 0, 1 \text{ or } 2, -\infty < t < \infty,$$

for all products leading to incompatible decisions not included in

$$\{d_i; i = 0, \dots, M-1\}$$

and is required in proceeding from the first to the second line of (4.17).

It then follows as before that the Bayes rule

$$(4.19) \quad \phi_{*}^{(n)}(t) = (\phi_{0*}^{(n)}(t) \dots \phi_{(M-1)*}^{(n)}(t)),$$

for the multiple comparisons problem is formed by the products

$$(4.20) \quad \phi_{i*}^{(n)}(t) = \prod_{pq \in N} \phi_{j_{pq}^i}^{pq}(t), \quad j_{pq}^i = 0, 1 \text{ or } 2,$$

of the elements of the Bayes rules  $\phi_*^{pq}(\mathbf{t})$  minimizing

$$(4.21) \quad A(\xi_n, \phi_*^{pq}(\mathbf{t})), \quad pq \in N,$$

provided these are compatible.

The first step in deriving  $\phi_*^{pq}(\mathbf{t})$  is practically identical with that ((3.12) to (3.14)) in Section 3. Thus

$$(4.22) \quad \phi_*^{pq}(\mathbf{t}) = (\phi_{0*}^{pq+}(\mathbf{t})\phi_{0*}^{pq-}(\mathbf{t}) \quad \phi_{1*}^{pq+}(\mathbf{t})\phi_{1*}^{pq-}(\mathbf{t}) \quad \phi_{0*}^{pq+}(\mathbf{t})\phi_{1*}^{pq-}(\mathbf{t}))$$

where, dropping the superscripts  $pq+$  or  $pq-$ ,  $\phi_*(\mathbf{t}) = (\phi_{0*}(\mathbf{t}) \quad \phi_{1*}(\mathbf{t}))$  now minimizes a subcomponent average risk of the form

$$(4.23) \quad A(\xi_n, \phi(\mathbf{t})) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{i=0}^1 L_i(\tau_{pq}) \phi_i(\mathbf{t}) f_n(\mathbf{t} | \boldsymbol{\tau}) d\mathbf{t} \xi_n(\boldsymbol{\tau}) d\boldsymbol{\tau}.$$

The elements of  $\boldsymbol{\tau}$  may be chosen so that  $\tau_{pq}$  is the first element  $1_1$  of  $\boldsymbol{\tau}$ . The compatibility condition

$$(4.24) \quad \phi_{1*}^{pq+}(\mathbf{t})\phi_{1*}^{pq-}(\mathbf{t}) = 0$$

must again be met.

The work of minimizing (4.23) follows closely that of minimizing (2.6) in Section 2 except that now the sample and parameter spaces have additional dimensions,  $(n-2)$  each. Dropping the subscript  $pq$  from  $\tau_{pq}$  the steps follow through with obvious changes till we get to

$$(4.25) \quad h_3^{(n)}(\tau, \mathbf{t}, v) \propto \tau \int_{-\infty}^{\infty} \int_0^{\infty} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n-1} [(ut_i - \tau_i)^2 + \tau_i^2/\gamma^2] \right\} u^{n-1} \psi(u|v) du d\tau_2$$

where the first integration is with respect to  $\tau_2$  defined as the last  $n-2$  elements of  $\boldsymbol{\tau} = (t \ \tau_2')'$ . This appears in place of  $h_2(\tau, t)$  as in (2.14) before which may now be denoted by  $h_3^{(3)}(\tau, \mathbf{t}, v)$ . On integrating with respect to  $\tau_2$  we get

$$(4.26) \quad \begin{aligned} h_3^{(n)}(\tau, \mathbf{t}, v) &\propto \tau \int_0^{\infty} \exp \left\{ -\frac{1}{2} \left[ (ut - \tau)^2 + \tau^2/\gamma^2 \right. \right. \\ &\quad \left. \left. - \sum_{i=2}^{n-1} u^2 t_i^2 / (1 + \gamma^2) \right] \right\} u^{n-1} u^{\nu-1} e^{-\frac{1}{2} v u^2} du \\ &\propto \tau \int_0^{\infty} \exp \left\{ -\frac{1}{2} [(u't' - \tau)^2 + \tau^2/\gamma^2] \right\} u' u'^{\nu'-1} e^{-\frac{1}{2} v' u'^2} du', \end{aligned}$$

where  $t = t_1$  is the first element of  $\boldsymbol{\tau}$  and is thus  $t = t_{pq}$ ,

$$(4.27) \quad u' = u/R, \quad t' = Rt, \quad R^2 = v' / \left[ v + \sum_{i=2}^{n-1} t_i^2 / (1 + \gamma^2) \right].$$

and  $v' = v + n - 2$ . Thus

$$(4.28) \quad h_3^{(n)}(\tau, t, v) \propto \tau \int_0^\infty \exp \left\{ -\frac{1}{2}[(ut' - \tau)^2 + \tau^2/\gamma^2] \right\} u \psi(u|v') du \\ = h_3^{(2)}(\tau, t', v').$$

Making a direct application of the remaining derivation in Section 2 it now follows that

$$(4.29) \quad \phi_*(t) = \begin{cases} (1, 0), & (t' < t_*) = (t < t_*/R), \\ (0, 1), & (t' > t_*) = (t > t_*/R) \end{cases}$$

where  $t_* = t_*(k, v', \gamma^2)$  is the same significant  $t$  value as before except that its degrees of freedom are now  $v' = v + n - 2$ .

Since  $t_*/R > 0$ , the compatibility condition (4.24) is met and applying (4.22) the component average risk (4.21) is minimized by

$$(4.30) \quad \phi_*^{pq}(t) = \begin{cases} (1 \ 0 \ 0), & (t'_{pq} < t_*) = (t_{pq} < t_*/R), \\ (0 \ 1 \ 0), & (t'_{pq} > t_*) = (t_{pq} > t_*/R), \\ (0 \ 0 \ 1), & (t'_{pq} < -t_*) = (t_{pq} < -t_*/R), \end{cases}$$

where  $t'_{pq} = Rt_{pq}$  for all  $pq \in N$ . Again since  $t_*/R > 0$ , the preceding compatibility conditions (4.18) are met and applying (4.19) the Bayes rule for the multiple comparisons problem is given by the simultaneous application (4.20) of all  $N = n(n-1)/2$  of the three-decision Bayes rules (4.30).

EXAMPLE 3. Suppose that  $n$  samples of yields like those of Example 2 have been obtained for  $n$  new treatments. For each and every pair  $(a, b)$  of treatments it is required to decide whether  $a$  can be recommended as the superior, whether  $b$  can or whether to withhold recommendations on both. Losses are scorable with respect to each pair of treatments as in Example 2, the loss ratio  $c$  being the same for all pairs, and are additive in giving the losses for each of the joint decisions to which they contribute. Risks are to be averaged with respect to a normal independent prior density for each of the means  $\mu_1, \dots, \mu_n$  each with the same mean and same variance  $\gamma^2\sigma^2/r$ . An invariant rule is required as before.

Because of sufficiency and invariance considerations the required rule can be restricted to depend on the observations through only the  $t$  vector

$$(4.31) \quad t = A\bar{x}/(s^2/r)^{1/2}$$

where  $A$  is as defined before for (4.5),  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)'$  is the vector of sample means and  $s^2$  is the pooled within-sample variance estimate

$$(4.32) \quad s^2 = \sum_{i=1}^n \sum_{j=1}^r (x_{ij} - \bar{x}_i)^2 / n(r-1)$$

with  $v = n(r-1)$  degrees of freedom.

The required decision rule is then given by the simultaneous application (4.20) of (4.30) where  $t_* = t_*(k, v', \gamma^2)$  with  $k = c-1$  and  $v' = v + n - 2 =$

$n(r-1) + n - 2$  and where  $t'_{pq}$  can be obtained by analysis-of-variance type steps as follow: Put  $S_t$ ,  $S_{pq}$ ,  $S'_{pq}$  and  $S_e$  for the treatment sum of squares, the sum of squares for the  $pq$  difference, the residual sum of squares for the  $pq$  difference and the error sum of squares

$$(4.33) \quad \begin{aligned} S_t &= r \sum_{i=1}^n \bar{x}_i - C, & S_{pq} &= r(\bar{x}_p - \bar{x}_q)^2/2, \\ S'_{pq} &= S_t - S_{pq}, & S_e &= \left( \sum_{i=1}^n \sum_{j=1}^r x_{ij}^2 - C \right) - S_t, \end{aligned}$$

respectively, where  $C$  is the correction term  $(\sum_{i=1}^n \sum_{j=1}^r x_{ij})^2/nr$ . Let  $s'^2_{pq}$  denote the pooled estimate of  $\sigma^2$  obtained as

$$(4.34) \quad s'^2_{pq} = [S_e + S'_{pq}/(1 + \gamma^2)]/v'.$$

Then  $t'_{pq}$  may be obtained as the square root of the variance ratio

$$(4.35) \quad t'^2_{pq} = S_{pq}/s'^2_{pq},$$

$t'_{pq}$  is given the same sign as  $\bar{x}_p - \bar{x}_q$ .

A more convenient rule for application can be obtained by expressing the inequalities  $t'^2_{pq} \leq t^2_*$  in the form  $d^2_{pq} \geq d^2_*$  where  $d_*$  is a least significant value for the difference  $d_{pq} = \bar{x}_p - \bar{x}_q$ . From  $t'^2_{pq} = t^2_*$  we get

$$(4.36) \quad \begin{aligned} t^2_* &= S_{pq}/s'^2_{pq} = v' S_{pq}/[S_e + (S_t - S_{pq})/(1 + \gamma^2)] \\ t^2_*[S_e + (S_t - S_{pq})/(1 + \gamma^2)] &= v' S_{pq} \\ S_{pq}[v' + t^2_*/(1 + \gamma^2)] &= t^2_*[S_e + S_t/(1 + \gamma^2)]. \end{aligned}$$

But  $S_{pq} = \frac{1}{2} r d^2_{pq}$ , hence this gives  $d^2_{pq} = d^2_*$  where

$$(4.37) \quad d_* = \left\{ \frac{2}{r} t^2_* [S_e + S_t/(1 + \gamma^2)] / [v' + t^2_*/(1 + \gamma^2)] \right\}^{\frac{1}{2}}$$

From this and a check on signs it follows that the multiple comparisons Bayes rule is given by the simultaneous application of the rules

$$(4.38) \quad \Phi^{pq}(\mathbf{t}) = \begin{cases} (1 \ 0 \ 0), & |d_{pq}| < d_*, \\ (0 \ 1 \ 0), & d_{pq} > d_*, \\ (0 \ 0 \ 1), & d_{pq} < -d_*. \end{cases}$$

## 5. Discussion.

**5.1. On the additional error degrees of freedom.** The emergence of  $t'_{pq} = d_{pq}/(2rs'^2)^{\frac{1}{2}}$  with  $v' = v + n - 2$  degrees of freedom as the component test statistic in the multiple comparisons solution may be surprising at first but less so after due consideration. In giving  $\mu_1, \dots, \mu_n$  identical independent normal distributions with variance  $\gamma^2 \sigma^2/r$  for risk-weighting purposes, the residual sum of squares between treatments,  $S_{pq} = r \sum_{i=2}^{n-1} y_i^2 = rs^2 \sum_{i=2}^{n-1} t_i^2$  in Example 3 for instance, is given the distribution of  $(1 + \gamma^2) \sigma^2 \chi^2_{n-2}$ . On this basis

$$(5.1) \quad s'^2_{pq} = [S_e + S_{pq}/(1 + \gamma^2)]/v'$$



becomes the appropriate estimator for  $\sigma^2$  in place of  $s^2 = S_e/v$  and the result is as might be expected. In practice a user might sometimes be reluctant to depend on the prior distribution assumption to this extent however, and might even want, see Subsection 5.3, to use  $S'_{pq}$ , or better  $S_t$ , to decide on an appropriate value  $\hat{\gamma}^2$  for  $\gamma^2$ . In such cases it would seem good sense to use a modified rule based on  $t_{pq}$  instead of  $t'_{pq}$ . This would consist of simultaneous applications of

$$(5.2) \quad \Phi_*^{pq}(\mathbf{t}) = \begin{cases} (1 \ 0 \ 0), & d_{pq} < d_*, \\ (0 \ 1 \ 0), & d_{pq} > d_*, \\ (0 \ 0 \ 1), & d_{pq} < -d_*, \end{cases}$$

where  $d_*$  is the least significant difference

$$(5.3) \quad d_* = (2s^2/r)^{1/2} t_*,$$

with  $t_* = t_*(k, v, \hat{\gamma}^2)$  based on  $v$  degrees of freedom.

5.2. *On the independence of the least significant difference and  $n$ .* By far the most striking feature of the multiple comparisons Bayes rule is the practically complete lack of dependence of the least significant difference  $d_*$  on  $n$ , the number of means involved. (The dependence of  $d_*$  on  $n$  via the estimation of error as discussed in the previous subsection is relatively trivial in this context and decreases to zero as  $v$  increases to  $\infty$ .) This is a direct consequence of the additive losses assumption similar results of which have also been treated by Duncan [3] and Thompson [15] and, in a more general form and context, by Lehmann [7]. In the past, a rule of this type, with  $d_*$  not increasing with  $n$ , has been considered more or less unacceptable. The main basis of objection has been the rapid increase in its so-called  $n$ -treatment significance level (Duncan [1] and [3]) or its experimentwise error rate (Tukey [17] and [18]),

$$(5.4) \quad \alpha_n = P[\text{rejecting } H_n \mid H_n], \quad H_n = H_n: \mu_1 \cdots = \mu_n,$$

with respect to  $n$ .

To illustrate, a non-increasing least significant difference of

$$d_* = 1.960 \sqrt{2}\sigma_p = 2.77\sigma_p$$

in the case  $v = \infty$  gives the experiment-wise error rates (found as upper-tail probabilities  $P[q_n > 2.77]$  of the range  $q_n$ )

$$(5.5) \quad \alpha_2 = .0500, \quad \alpha_3 = .1223, \quad \alpha_4 = .2034, \cdots, \quad \alpha_{20} = .9183.$$

The possibility in this case of wrongly rejecting the homogeneity hypothesis for 20 means, for example, with a probability of 91.83 per cent, may at first appear to be unacceptably high. As a result, procedures have been proposed with increasing significant differences aimed at suppressing the rapid increases in  $\alpha_n$ . These have varied considerably from rapidly increasing significant differences such as (in comparable cases, dropping the factor  $\sigma_p$ )

$$(5.6) \quad 2.77, 3.32, 3.63, \cdots, 5.01,$$

to more slowly increasing ones such as

$$(5.7) \quad 2.77, 2.92, 3.02, \cdots, 3.47,$$

depending on the relative importance attached to experimentwise error rates and the degree to which they should be suppressed. The differences (5.6) termed honest significant differences by Tukey [17] [18] suppress all of the experimentwise error rates to .0500. The differences (5.7) proposed by the author [1] and [3] suppress them to less conservative so-called levels based on degrees of freedom

$$(5.8) \quad \alpha'_2 = .0500, \quad \alpha'_3 = .0975, \quad \alpha'_4 = .1426, \dots, \quad \alpha'_{20} = .6415,$$

obtained as  $\alpha'_n = 1 - (1 - \alpha_2)^{n-1}$ . A less conservative procedure yet is the one due to Fisher [5] that uses the same least significant difference for all  $n$  (2.77 in the case above) provided that first the homogeneity hypothesis  $H_n$  can be rejected by an  $F$  ratio test.

Now it appears that, in a Bayes sense, provided the losses are additive and other things (e.g.,  $k$  and  $\gamma^2$ ) are equal, the same least significant difference is optimum no matter how large the number  $n$  of treatments involved. (From Lehmann's work [7] it is clear that this would also apply under other optimality criteria such as, for example, minimax.) The high experimentwise error-rate of 91.83 per cent in the case quoted might well be worth tolerating for instance, because, it might be said, of the relatively low prior probability of the hypothesis  $H_{20}$  involved and its relative unimportance among so many others.

The inverse form of dependence of the least significant difference  $d_*$  on the risk-weighting variance ratio  $\gamma^2$  may do much to reconcile its independence of  $n$  with at least some of the common almost instinctive urge to make it increase with  $n$ . In the case  $v = \infty$  it is directly proportional to  $(1 + 1/\gamma^2)^{1/2}$  and approximately so for smaller values of  $v$ . If, in the conduct of a large experiment, the treatments under study have a lower anticipated heterogeneity than those which would have been studied in a more limited experiment with fewer populations, a lower risk-weighting variance ratio  $\gamma^2$  would be appropriate and hence would be a larger least significant difference. Such a situation could often arise in practice, and, if  $\gamma^2$  is varying in an interval of small values this could make a substantial increase in the significant difference with an increase in  $n$ . On the other hand, however, the reverse situation could also arise. In selection experiments, for example, the treatments under consideration may be the top  $n$  performers as assessed by experiences in previous trials. Here, the larger the number of treatments it has been possible to include in an experiment, the larger will be the appropriate  $\gamma^2$  and hence, the smaller the least significant difference.

*5.3. A practical adaptation of the Bayes rule.* In the complete absence of prior criteria for choosing  $\gamma^2$ , the user might sensibly, it would seem, obtain an estimate of it from the variance ratio

$$(5.9) \quad F = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{y_i^2}{s_i^2} \quad \left( = \frac{1}{n-1} \frac{S_t}{s^2} \text{ in Example 3} \right),$$

employed in the preliminary  $F$  test of Fisher's least-significant-difference procedure. Since the ratio of the corresponding expectations is  $1 + \gamma^2$  he might for example put  $\gamma^2 = F - 1$ , enter Table 1 with  $\gamma^2 = \hat{\gamma}^2$ , and use the simple direct rule as given (5.2) in Subsection 5.1. It is of considerable interest to note the

closeness of Fisher's earlier procedure to this type of adaptation. Both rules depend on a preliminary inspection of the  $F$  ratio. In Fisher's rule a big or small  $F$  ratio leads to the use of an independently chosen least significant difference on a go-no basis. In the new rule a big or small  $F$  ratio leads to the use of a small or big least significant difference on a continuously related basis.

**5.4 Concluding remarks.** As presented the model is limited to a class of symmetric problems in which the loss and prior probabilities are invariant with respect to all  $n!$  permutations of the means involved. Many problems in practice however are naturally taken to be symmetric in this way. Within this class, the assumptions of linear losses and normal prior densities would seem directly appropriate for some problems and useful at least as good approximations for many others.

From the given development it is fairly clear that similar rules for a wider class of less symmetrical problems can be obtained leading to the use of different significant  $t$  ratios for each of the component or even each of the subcomponent problems involved. Development and discussion of these are deferred to a further paper.

A most interesting point is one raised (in private correspondence) by Professor F. J. Anscombe following from the type of discussion in Subsection 5.3. In addition to providing an estimate of the prior variance  $\gamma^2$  and therefore a means of rejecting an assumed value for this parameter, the data may provide evidence for reasonably rejecting other assumed features of the prior density. Further developments are needed for handling problems of this type.

In conclusion, it is worth repeating, the most important result discussed in Subsection 5.2 namely the independence of the least-significant-difference  $d_*$  and the number of components problems involved, depends only on the additivity assumption for the losses. It is independent of the form of the component loss functions and of prior density assumed. It appears further that the same result would follow even if the class of component problems were extended to include all contrasts among the means as considered by Scheffé [11]. Thus in a symmetric situation, for example, the same significant  $t$  ratio would be appropriate whether it be desired to test just one comparison chosen *a priori*, the set of all  $\frac{1}{2}n(n-1)$  comparisons in the multiple comparisons problem or the set of all contrasts. The additivity-of-losses assumption on which this critically depends appears to be a reasonable one, and appropriate to many practical situations.

**6. Computation of significant  $t$  ratios.** In computing the values in Table 1 the ratios  $g(y)/g(-y)$  in (2.22) may be simplified first to  $g_1(y)/g_1(-y)$  where

$$(6.1) \quad g_1(y) = \begin{cases} (1-y^2)^{\frac{1}{2}} + y \sin^{-1} y + \pi y/2, & v = 1, \\ (1+y)^2 & v = 2, \\ 2(1-y^2)^{\frac{1}{2}} + 3y^2(1-y^2)^{\frac{1}{2}} + 3y \sin^{-1} y + 3\pi y/2, & v = 3, \\ (1+y)^3/(3+y), & v = 4, \\ (1+y)^4/(5+4y+y^2), & v = 6, \\ (1+y)^5/(429+1384y+2063y^2+1776y^3 \\ \quad + 915y^4+264y^5+33y^6), & v = 14, \end{cases}$$

or to  $g_2(z)/g_2(-z)$  where

$$(6.2) \quad g_2(z) = f(z) + zF(z), \quad v = \infty,$$

and  $f$  and  $F$  are the standard normal density and cumulative distribution functions respectively. These follow readily from (2.23) except for (6.2) in the case  $v = \infty$  which can be obtained as follows.

Putting  $y = z/(v + z^2)^{1/2}$  and thus  $(1 - y^2) = 1/(1 + z^2/v)$  and  $dy = dz/[v^{1/2}(1 + z^2/v)^{3/2}]$  in  $g(y)$  in (2.23) we get

$$(6.3) \quad g(y) = \frac{1}{(1 + z^2/v)^{v/2}} + \frac{z}{(1 + z^2/v)^{1/2}} \int_0^z \frac{du}{(1 + u^2/v)^{(v+1)/2}} + \frac{(\pi v)^{1/2}[(v-2)/2]! z}{[(v-1)/2]! 2(1 + z^2/v)^{1/2}}.$$

Recalling that the probability density function of the Student  $t$  distribution is

$$(6.4) \quad h(t|v) = \frac{[(v-1)/2]!}{(\pi v)[(v-2)/2]!} \cdot \frac{1}{(1 + t^2/v)^{(v+1)/2}}$$

we may write

$$(6.5) \quad g(y) \propto (1 + z^2/v)^{1/2} h(z|v) + \frac{z[H(z|v) - \frac{1}{2}]}{(1 + z^2/v)^{1/2}} + \frac{z/2}{(1 + z^2/v)^{1/2}} \\ \propto (1 + z^2/v) h(z|v) + zH(z|v) = g_2(z) \text{ say,}$$

where  $H(z|v)$  is the cumulative distribution  $\int_{-\infty}^z h(t|v) dt$ . Treating  $g(-y)$  in the same way we reach the result

$$(6.6) \quad g(y)/g(-y) = g_2(z)/g_2(-z),$$

which reduces to  $g_2(z)/g_2(-z)$  as  $v \rightarrow \infty$ .

Next,  $y_*(k, v)$  or  $z_*(k)$  is found as the solution of  $g_1(y)/g_1(-y) = k$  or  $g_2(z)/g_2(-z) = k$  respectively. Finally  $t_*$  is found as the positive square root of

$$(6.7) \quad t_*^2 = v/(\beta^{-2} y_*^{-2} - 1) \quad \text{or from} \quad t_* = z_*/\beta.$$

**Acknowledgments.** Help from Professor Harold Hotelling and other colleagues is gratefully acknowledged.

#### REFERENCES

- [1] DAVID B. DUNCAN, "A significance test for differences between ranked treatments in an analysis of variance," *Virginia J. Sci.*, Vol. 2 (1951), pp. 171-189.
- [2] DAVID B. DUNCAN, "On the properties of the multiple comparisons test," *Virginia J. Sci.*, Vol. 3 (1952), pp. 50-67.
- [3] DAVID B. DUNCAN, "Multiple range and multiple  $F$  tests," *Biometrics*, Vol. 11 (1955), pp. 1-42.
- [4] WALTER T. FEDERER, *Experimental Design*, MacMillan Co., New York, 1955.
- [5] R. A. FISHER, *The Design of Experiments*, six eds. Oliver and Boyd, London (1935-1951).
- [6] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1-26.

- [7] E. L. LEHMANN, "A theory of some multiple decision problems, I," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1-25.
- [8] E. L. LEHMANN, "A theory of some multiple decision problems, II," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 547-572.
- [9] JEROME C. R. LI, *Introduction to Statistical Inference*, Edwards Bros., Ann Arbor, 1957.
- [10] D. NEWMAN, "The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of the standard deviation," *Biometrika*, Vol. 31 (1939), pp. 20-30.
- [11] HENRY SCHEFFÉ, "A method for judging all contrasts in the analysis of variance," *Biometrika*, Vol. 40 (1953), pp. 87-104.
- [12] HENRY SCHEFFÉ, *"The Analysis of Variance"*, John Wiley and Sons, New York, 1959.
- [13] G. W. SNEDECOR, *Statistical Methods*, 5th ed., Iowa State College Press, Ames, 1957.
- [14] R. G. D. STEEL AND J. H. TORRIE, *"Principles and Procedures of Statistics,"* McGraw-Hill Book Co., New York, 1960.
- [15] W. A. THOMPSON, "A note on ranking means," *Ann. Math. Stat.*, Abstract 13, Vol. 27 (1956).
- [16] J. W. TUKEY, "Comparing individual means in the analysis of variance," *Biometrics*, Vol. 5 (1949), pp. 99-114.
- [17] J. W. TUKEY, "Quick and dirty methods in statistics, Part II, Simple analyses for standard designs," *Proc. Fifth Ann. Convention, Amer. Soc. for Quality Control*, (1951), pp. 189-197.
- [18] J. W. TUKEY, "The problem of multiple comparisons," unpublished dittoed notes, 396 pp., Princeton University (1953).

# THE USE OF LEAST FAVORABLE DISTRIBUTIONS IN TESTING COMPOSITE HYPOTHESES

BY H. E. REINHARDT

Montana State University<sup>1</sup>

**0. Introduction and summary.** The usual method of finding a most powerful size  $\alpha$  test of a composite hypothesis against a simple alternative is the guessing of a Least Favorable Distribution (LFD)—introduced at various levels of generality by Neyman and Pearson [6], Wald [7], Lehmann [4], and Lehmann and Stein [5]—and testing the mixture of the distributions of the hypothesis over this LFD against the alternate using the Neyman-Pearson Fundamental Lemma. In guessing LFD's statisticians have looked for a mixture which is "like" the alternate.

In this paper, the notion of Uniformly Least Favorable Mixture (ULFM) is introduced. In Section 2, we show that a ULFM is a point in the convex set of mixtures of the hypothesis which is closest (in the sense of the  $\mathcal{L}^1$  norm) to the alternate. The condition is not sufficient. More generally, any LFM corresponds to a point which is closest to the alternate in some expansion or contraction of this set of mixtures. A sufficient condition for ULFM's is, essentially, that the nuisance parameter can take on the same values in the alternate as in the hypothesis. In Section 3, we consider the case where no ULFM exists. We show, *inter alia*, that any distribution is least favorable for a closed set of  $\alpha$ 's. (A pathological example shows that this closed set need not be the union of a finite number of closed intervals.)

**1. Notation and definitions.** We consider a family  $f_\theta$ ,  $\theta \in \Omega$ , of densities and a density  $g$  with respect to a  $\sigma$ -finite measure  $\mu$  over a measurable space  $(\mathcal{X}, \mathfrak{A})$ . For tests  $\varphi$ , i.e., measurable functions  $\varphi$  such that  $0 \leq \varphi(x) \leq 1$  for all  $x$ , we shall use the inner product notation

$$(1) \quad (\varphi, f_\theta) = \int_{\mathcal{X}} \varphi(x) f_\theta(x) d\mu(x).$$

For the problem of testing a composite hypothesis  $H: f_\theta$ ,  $\theta \in \Omega$ , against the simple alternative  $g$  a most powerful level  $\alpha$  test is a test  $\varphi$  which maximizes the power  $(\varphi, g)$  among all tests satisfying  $(\varphi, f_\theta) \leq \alpha$  for all  $\theta \in \Omega$ .

We assume there is a  $\sigma$ -algebra  $\mathfrak{B}$  on the indexing set  $\Omega$  such that  $f_\theta(x)$  is measurable on  $\mathfrak{A} \times \mathfrak{B}$ . If  $\lambda$  is a probability measure over  $(\Omega, \mathfrak{B})$  we define the mixture  $f_\lambda$  by

$$(2) \quad f_\lambda(x) = \int_{\Omega} f_\theta(x) d\lambda(\theta).$$

Received June 6, 1960; revised July 24, 1961.

<sup>1</sup> Address during the academic year 1961-62: Army Mathematics Research Center, Madison, Wisconsin.



Then  $\lambda$  is called a *least favorable distribution* and  $f_\lambda$  a *least favorable mixture* (LFM) at level  $\alpha$  for testing  $H: f_\theta, \theta \in \Omega$ , against  $g$  if a most powerful level  $\alpha$  test of  $f_\lambda$  against  $g$  is also most powerful for testing  $H$  against  $g$ . If  $f_\lambda$  is least favorable at every level  $\alpha, 0 \leq \alpha \leq 1$ ,  $f$  is called a *uniformly least favorable mixture* (ULFM) for testing  $H$  against  $g$ .

To obtain a most powerful test for testing  $H: f_\theta, \theta \in \Omega$ , against  $g$  and a least favorable distribution  $\lambda$  or mixture  $f_\lambda$  we shall frequently use the following generalization of the fundamental lemma of Neyman and Pearson by Lehmann and Stein ([5], Theorem 1, or Corollary 5, p. 92, of [3]):

**THEOREM 1.1.** *Suppose that  $\lambda$  is a probability distribution over  $\Omega$  and that  $\Omega'$  is a subset of  $\Omega$  with  $\lambda(\Omega') = 1$ . Let  $\varphi$  be a test such that*

$$\varphi(x) = 1 \quad \text{if } g(x) > kf_\lambda(x)$$

$$\varphi(x) = 0 \quad \text{if } g(x) < kf_\lambda(x).$$

*Then  $\varphi$  is a most powerful level  $\alpha$  test for testing  $H$  against  $g$  provided*

$$(3) \quad (\varphi, f_{\theta'}) = \sup_{\theta \in \Omega} (\varphi, f_\theta) = \alpha \quad \text{for all } \theta' \in \Omega'.$$

**2. Uniformly least favorable distributions.** In [3] and [5] Lehmann and Lehmann and Stein, respectively, give examples of problems where the least favorable distribution depends on the level of significance. The case where the LFD is independent of the level of significance is more tractable and we consider it first. The following theorem shows the relation between ULFM's and the  $\mathcal{L}^1$  norm:

**THEOREM 2.1.** *If  $f_\lambda$  is a ULFM for testing  $H: f_\theta, \theta \in \Omega$ , against  $g$ , then  $f_\lambda - g$  is a point of smallest norm in the convex set  $\{f, -g\}$ . (Here  $\{f, -g\}$  is the convex set of mixtures of  $f_\theta$ 's formed by averaging with respect to a probability measure on the space  $\Omega$ . Thus  $\{f, -g\}$  is a convex set in the positive part of the unit sphere in an  $\mathcal{L}^1$  space.)*

We omit the direct proof of Theorem 2.1 since it is a special case (for  $k = 1$ ) of Theorem 3.2.

For Bernoulli distributions with a single trial the situation is particularly simple. The positive part of the unit sphere is the line segment in the Euclidean plane consisting of points  $(x, 1 - x)$  with  $0 \leq x \leq 1$ . Thus the convex set spanned by the hypothesis is a line segment which we can take to be closed. The closest point to the alternate is the one whose first co-ordinate minimizes  $|x - p|$  (where  $(p, 1 - p)$  is the distribution under the alternate). This closest point is, in fact, a ULFM. (If  $\min |x - p| = 0$ , the alternate is in the hypothesis and  $\varphi(x) \equiv \alpha$  is most powerful. If  $\min |x - p| \neq 0$  there is a non-trivial test.)

If we were able to find the LFM for a sample of  $n$  independent observations knowing it for a sample of one observation, we would not have to do the problem over for each sample size. Such an inductive procedure is possible in a fairly special case.

LEMMA 2.2. If  $\mu\{x \mid f(x) = kg(x)\} = 0$  for every  $k$ , then

$$(\mu \times \mu)\{(x, y) \mid f(x)f(y) = kg(x)g(y)\} = 0$$

for every  $k$ .

PROOF. Let  $E = \{(x, y) \mid f(x)f(y) = kg(x)g(y)\}$ . Consider a section  $E^{y_0}$  of  $E$  by  $y_0$ , i.e., the set of all  $x$  such that  $(x, y_0) \in E$ . Hence  $x \in E^{y_0}$  if and only if

$$f(x) = kg(x)g(y_0)/f(y_0).$$

Then by the condition of the lemma  $\mu(E^{y_0}) = 0$ , and this is true for all  $y_0$  except possibly the set where  $f(y_0) = 0$ . But this set has measure 0 by condition (taking  $k = 0$ ). Thus all  $y$  sections have measure 0 and hence  $E$  has product measure 0 ([2], Theorem 36A).

THEOREM 2.3. Let  $f_{\theta_0}$  be a ULFM for testing  $H_1: f_{\theta}$ ,  $\theta \in \Omega$ , against  $g$ ; if  $\theta_0 \in \Omega$  and if  $\mu\{x \mid f_{\theta_0}(x) = kg(x)\} = 0$  for each  $k$ , then for  $n$  independent observations of  $X$ , the density  $\prod_{i=1}^n f_{\theta_0}(x_i)$  is uniformly least favorable for testing  $H_n: \prod_{i=1}^n f_{\theta_0}(x_i)$ ,  $\theta \in \Omega$ , against  $\prod_{i=1}^n g(x_i)$ .

PROOF. We prove the statement only for  $n = 2$ . The proof can easily be generalized by an induction.

By Lemma 2.2 and the Neyman-Pearson Lemma there is for each  $\alpha$  a most powerful level  $\alpha$  test  $\varphi$  for testing  $f_{\theta_0}(x)f_{\theta_0}(y)$  against  $g(x)g(y)$  such that

$$\varphi(x, y) = 1 \quad \text{if} \quad g(x)g(y) > kf_{\theta_0}(x)f_{\theta_0}(y), \quad \text{and}$$

$$\varphi(x, y) = 0 \quad \text{if} \quad g(x)g(y) \leq kf_{\theta_0}(x)f_{\theta_0}(y).$$

Thus  $\varphi$  is the indicator function  $I_S$  of a set  $S$ . Since  $f_{\theta_0}$  is a ULFM, the section  $S^y$  of  $S$  by  $y$  is a most powerful test for testing  $H_1$  against  $g$  for  $g(y) > 0$  and  $S^y = \emptyset$  for  $g(y) = 0$ . Hence

$$(4) \quad \int_{S^y} [f_{\theta_0}(x) - f_{\theta}(x)] d\mu(x) \geq 0$$

and similarly

$$(5) \quad \int_{S^x} [f_{\theta_0}(y) - f_{\theta}(y)] d\mu(y) \geq 0.$$

Applying Fubini's theorem we obtain

$$(6) \quad \int_S [f_{\theta_0}(x)f_{\theta_0}(y) - f_{\theta}(x)f_{\theta}(y)] d(\mu \times \mu) \geq 0.$$

Thus  $\varphi = I_S$  is uniformly most powerful for testing  $H_2$  against  $g(x)g(y)$  ([3], Theorem 3.7). Hence  $f_{\theta_0}(x)f_{\theta_0}(y)$  is a ULFM.

The LFM has to be in  $H_1$ , for otherwise we would be looking at

$$\prod_{i=1}^n \int f_{\theta}(x_i) d\lambda(\theta);$$

and this, in general, is not of the form  $\int \prod_{i=1}^n f_{\theta}(x_i) d\lambda(\theta)$ . It is mixtures of the latter sort that are available as potential LFM's.

In many densities of practical importance, the most natural parametrization is given by an indexing set which is a product space. (For normal distributions the pair  $(\mu, \sigma^2)$  where  $\mu$  is the mean and  $\sigma^2$  the variance is a natural choice.) In many hypothesis testing problems, the statistician is only interested in one co-ordinate of the parameter. However, the nature of the observations he can make forces him to consider the other co-ordinates (traditionally called "nuisance parameters").

For problems of a certain form we can eliminate the nuisance parameters from consideration. Toward this end, let  $X, Y$  be independent random variables whose joint probability measure is absolutely continuous with respect to the product measure  $\mu \times \nu$ ; thus,

$$(7) \quad P((X, Y) \in A) = \iint_A f_{\theta}(x) g_{\eta}(y) d\mu \times \nu.$$

**THEOREM 2.4.** *If there exists a probability measure  $\lambda$  on  $H$  such that  $h(y) = \int_H g_{\eta}(y) d\lambda(\eta)$ , then  $f_{\theta_0}(x)h(y)$  is a ULFM for testing  $H: f_{\theta_0}(x)g_{\eta}(y), \eta \in H$ , against  $f_{\theta_1}(x)h(y)$ .*

**PROOF.** A most powerful level  $\alpha$  test for testing  $f_{\theta_0}(x)h(y)$  against  $f_{\theta_1}(x)h(y)$  is given by

$$\begin{aligned} \varphi(x) &= 1 & \text{if } f_{\theta_1}(x)h(y) > kf_{\theta_0}(x)h(y) \\ \varphi(x) &= 0 & \text{if } f_{\theta_1}(x)h(y) < kf_{\theta_0}(x)h(y), \end{aligned}$$

where  $k$  is chosen so  $\int \varphi(x)f_{\theta_0}(x)d\mu = \alpha$ . Since  $\varphi$  is independent of  $y$  it follows that  $\varphi$  is most powerful for testing  $H$  against  $f_{\theta_1}(x)h(y)$ . Since  $\alpha$  is arbitrary  $f_{\theta_0}(x)h(y)$  is uniformly least favorable. (Theorem 2.1 suggests that this is a natural candidate for ULFM.)

**COROLLARY 2.5.** (See [1], p. 86.) *For testing  $H: f_{\theta_0}(x)g_{\eta}(y), \eta \in H$ , against  $f_{\theta_1}(x)g_{\eta}(y), \eta \in H$ , there is a uniformly most powerful test given by*

$$\begin{aligned} \varphi(x, y) &= 1 & \text{if } f_{\theta_1}(x) > kf_{\theta_0}(x) \\ \varphi(x, y) &= 0 & \text{if } f_{\theta_1}(x) < kf_{\theta_0}(x). \end{aligned}$$

**PROOF.** Apply the theorem to the problem of testing  $f_{\theta_0}(x)g_{\eta_1}(y), \eta \in H$ , against  $f_{\theta_1}(x)g_{\eta_1}(y)$  for each  $\eta_1$ . A ULFM is  $f_{\theta_0}(x)g_{\eta_1}(y)$ .

The corollary says, in effect, that we can not get information about  $\theta$  by observing a random variable whose distribution is independent of  $\theta$ !

Theorem 2.4 is a generalization of the method to obtain a uniformly most powerful test for the following example, used by Lehmann and Stein [5] (or see [3], p. 96).

**EXAMPLE.** Let  $X_1, \dots, X_n$  be independently normally distributed with mean  $\eta$  and variance  $\theta$ . For given  $\theta_1 > \theta_0$ , we wish to test the composite hypothesis  $H: \theta = \theta_0, -\infty < \eta < +\infty$ , against the simple alternative  $\theta = \theta_1, \eta = \eta_1$ .

The normal distribution of the parameter  $\eta$  with mean  $\eta_1$  and variance  $(\theta_1 - \theta_0)/n$  is uniformly least favorable.

**3. The non-uniform case.** In the case where no ULFM exists, a result similar to Theorem 2.1 holds.

LEMMA 3.1. Let  $f$  and  $g$  be probability densities with respect to  $\mu$ ; if  $k > 0$ , then

$$(8) \quad \|kf - g\| = k - 1 + 2 \sup_{A \in \mathfrak{B}} \int_A (g(x) - kf(x)) d\mu.$$

PROOF.

$$\begin{aligned} \|kf - g\| &= \int_{\mathfrak{X}} |(kf(x) - g(x))| d\mu \\ &= \int_{\mathfrak{X}} (kf(x) - g(x)) d\mu + 2 \int_{g > kf} (g(x) - kf(x)) d\mu \\ &= k - 1 + 2 \sup_{A \in \mathfrak{B}} \int_A (g(x) - kf(x)) d\mu. \end{aligned}$$

THEOREM 3.2. Suppose  $f_{\lambda\alpha}(x)$  is an LFM at the level  $\alpha$  for testing  $H: f_\theta, \theta \in \Omega$ , against  $g$  and the most powerful test is given by

$$\begin{aligned} \varphi(x) &= 1 \quad \text{if } g(x) > kf_{\lambda\alpha}(x), \\ \varphi(x) &= 0 \quad \text{if } g(x) < kf_{\lambda\alpha}(x), \end{aligned}$$

then  $kf_{\lambda\alpha} - g$  is a point of smallest norm in the convex set  $\{kf_\theta - g\}$ .

PROOF. Consider any mixture  $f_\theta$  of  $f_\theta$ 's. Since  $\varphi(x)$  is a most powerful level  $\alpha$  test we have

$$(9) \quad \int \varphi(x) f_\theta(x) d\mu \leq \alpha.$$

Hence by Lemma 3.1 and by definition of  $\varphi$

$$\begin{aligned} \|kf_\theta - g\| &= k - 1 + 2 \sup_{A \in \mathfrak{B}} \int_A (g(x) - kf_\theta(x)) d\mu \\ &\geq k - 1 + 2 \int \varphi(x) (g(x) - kf_\theta(x)) d\mu \geq \|kf_{\lambda\alpha} - g\|. \end{aligned}$$

In solving a problem where there is no ULFM, we would like to know how to proceed from an LFM for  $\alpha_1$  to one for  $\alpha_2$ . We prove three useful theorems.

THEOREM 3.3. The set of LFM's for a particular  $\alpha$  is convex.

PROOF. Suppose  $f_1$  and  $f_2$  are both LFM's at level  $\alpha$  and suppose  $0 \leq a \leq 1$ . Let a most powerful test be  $\varphi$ . Then

$$\begin{aligned} \varphi(x) &= 1 \quad \text{if } g(x) > kf_1(x) \\ \varphi(x) &= 0 \quad \text{if } g(x) < kf_1(x) \end{aligned}$$

for  $i = 1, 2$ . Let  $k_0$  and  $b$  be determined by

$$(10) \quad k_0 a = k_1 b \quad \text{and} \quad k_0(1 - a) = k_2(1 - b).$$

Then

$$\varphi(x) = 1 \quad \text{if} \quad g(x) > k_0[af_1(x) + (1 - a)f_2(x)]$$

$$\varphi(x) = 0 \quad \text{if} \quad g(x) < k_0[af_1(x) + (1 - a)f_2(x)].$$

Thus  $\varphi$  is a most powerful test for  $af_1 + (1 - a)f_2$  vs.  $g$  and consequently  $af_1 + (1 - a)f_2$  is least favorable.

COROLLARY 3.4. Suppose  $f_1$  and  $f_2$  are both least favorable at level  $\alpha$ , and the  $k$ 's of the fundamental lemma are  $k_1$  and  $k_2$  with  $k_1 \geq k_2 > 0$ . Then for any  $k_0$ ,  $k_1 \geq k_0 \geq k_2$ , there is an  $a$ ,  $0 \leq a \leq 1$ , such that the  $k$  of the fundamental lemma for testing  $af_1 + (1 - a)f_2$  against  $g$  is  $k_0$ .

PROOF. Determine  $a$  and  $b$  from (10). The corollary then follows from the proof of Theorem 3.3.

THEOREM 3.5. For testing  $H: f_\theta$ ,  $\theta \in \Omega$ , against  $g$ , any  $f_\lambda$  is least favorable for a closed set of  $\alpha$ 's (which may be empty).

PROOF. Suppose  $\lim_n \alpha_n = \alpha$  and  $f_\lambda$  is least favorable at level  $\alpha_n$  for each positive integer  $n$ . We show  $f_\lambda$  is least favorable at level  $\alpha$ . Let the most powerful level  $\alpha_n$  test be  $\varphi_n$ . Then, because of the weak\* compactness of the set of tests, ([3], Appendix 4) there is a  $\varphi(x)$ ,  $0 \leq \varphi(x) \leq 1$  and a subsequence of  $\{\varphi_n\}$ —which we take to be  $\{\varphi_n\}$  itself such that  $\lim_n (\varphi_n, f) = (\varphi, f)$  for every  $f$ . In particular we have (since  $\varphi_n$  is level  $\alpha_n$ )

$$(11) \quad (\varphi_n, f_\theta) \leq \alpha_n$$

so that

$$(12) \quad (\varphi, f_\theta) \leq \alpha.$$

Thus  $\varphi$  is level  $\alpha$  for  $H$ .

Similarly, since  $f_\lambda$  is least favorable at level  $\alpha_n$ ,

$$(13) \quad (\varphi, f_\lambda) = \alpha$$

and  $\varphi$  is level  $\alpha$  for  $f_\lambda$  vs.  $g$ .

Let the power of  $\varphi_n$  be  $1 - \beta_n$ , so

$$(14) \quad (\varphi_n, g) = 1 - \beta_n.$$

Now power is a concave and hence continuous function of size so the most powerful level  $\alpha$  test of  $f_\lambda$  vs.  $g$  has power  $1 - \beta = \lim_n (1 - \beta_n)$ . But  $(\varphi, g) = 1 - \beta$ , so  $\varphi$  is a most powerful level  $\alpha$  test of  $f_\lambda$  vs.  $g$ , and satisfies the size requirements of the original problem. Hence  $f_\lambda$  is least favorable.

Before proceeding to the final theorem of this section, we remark that for testing  $f$  vs.  $g$  (both simple) and for any  $k > 0$ , there is an  $\alpha$  such that the most powerful level  $\alpha$  test of  $f$  vs.  $g$  is

$$\begin{aligned}\varphi(x) &= 1 && \text{if } g(x) > kf(x) \\ \varphi(x) &= 0 && \text{if } g(x) < kf(x).\end{aligned}$$

Clearly we can take any  $\alpha$  satisfying

$$(15) \quad \int_{kf < g} f d\mu \leq \alpha \leq \int_{kf \leq g} f d\mu.$$

Further, if we let  $k_\alpha$  be the (not necessarily unique)  $k$  of the Fundamental Lemma such that  $\varphi$  is a most powerful level  $\alpha$  test of  $f$  vs.  $g$ , then  $\alpha_2 < \alpha_1$  implies  $k_{\alpha_1} \leq k_{\alpha_2}$ .

**THEOREM 3.6.** *Suppose for testing  $H: f_\theta, \theta \in \Omega$ , against  $g$  there are a finite number of mixtures of the hypothesis,  $f_1, \dots, f_m$ , (not necessarily distinct) each least favorable for an interval  $I_i$  of  $\alpha$ 's such that*

$$(16) \quad \bigcup_{i=1}^m I_i = [0, 1].$$

*Then there is an LFM for some  $\alpha$  which is a point closest to  $g$  (in  $\mathcal{L}^1$  norm) among the points of the convex set  $\{f_i\}$ .*

**PROOF.** Let the ends of the intervals be  $0 = \alpha_0 < \alpha_1 < \dots < \alpha_m = 1$ . We can assume, without loss of generality, that  $f_i$  is least favorable for  $\alpha_{i-1} \leq \alpha \leq \alpha_i$ . Let  $k(\alpha, i)$  be the  $k$  of the Fundamental Lemma for testing  $f_i$  against  $g$  at level  $\alpha$ . We can take  $k(\alpha_0, 1) = \infty$  and  $k(\alpha_m, m) = 0$ . Hence there is an  $i$  such that  $k(\alpha_{i-1}, i) \geq 1 \geq k(\alpha_i, i)$  or  $k(\alpha_i, i) \geq 1 \geq k(\alpha_i, i+1)$ . In the first case, by the remark above there is an  $\alpha, \alpha_{i-1} \leq \alpha \leq \alpha_i$ , such that  $k(\alpha, i) = 1$  and hence by Theorem 3.1  $f_i$  is closest to  $g$ . In the second case by Theorem 3.3,  $f_\lambda = af_i + (1-a)f_{i+1}$  is least favorable at level  $\alpha_i$ , and for a suitable choice of  $a$  the  $k$  of the Fundamental Lemma is 1 for  $f_\lambda$  by Corollary 3.4. Then  $f_\lambda$  is a point closest to the alternate.

Thus one might be able to proceed stepwise—finding an LFM as a mixture closest to the alternate, then finding the set of  $\alpha$ 's for which this is least favorable. For each point which is a boundary point of this set of  $\alpha$ 's there may be another LFM. (This is true if there are only a finite number of LFM's.) For this LFM, we could proceed as with the first one. This procedure works perfectly for some problems which will be considered elsewhere. The theorems suggest, however, that LFM's are going to be difficult to find in general.

We conclude with a pathological example. We take as sample space the positive integers; the distributions can be represented as sequences  $\{a_n\}$  with  $a_n \geq 0$  and  $\sum a_n = 1$ . We consider three sequences defined inductively.

$$\begin{array}{lll} a_1 = 1/3 & b_1 = 1/3 & c_n = (1/2)^n \\ a_2 = 2/9 + \epsilon & b_2 = 2/9 - \epsilon & \\ a_3 = 4/27 - \epsilon & b_3 = 4/27 + \epsilon & \\ a_4 = 2^3/3^4 & b_4 = 2^3/3^4 & \end{array}$$



$$\begin{aligned} a_5 &= (2/3)^3 b_2 & b_5 &= (2/3)^5 a_2 \\ a_6 &= (2/3)^3 b_3 & b_6 &= (2/3)^3 a_3 \\ &\vdots & & \\ a_{n+\epsilon} &= (2/3)^6 a_n & b_{n+\epsilon} &= (2/3)^6 b_n \end{aligned}$$

where  $0 < \epsilon < 2/81$ . Then the sequences  $\{a_n/c_n\}$  and  $\{b_n/c_n\}$  are monotone increasing and

$$\sum_{n=1}^{6k+2} a_n > \sum_{n=1}^{6k+2} b_n \quad \text{and} \quad \sum_{n=1}^{6k+5} b_n > \sum_{n=1}^{6k+5} a_n \quad \text{for } k = 0, 1, 2, \dots$$

Hence for testing  $H: \{a_n\}, \{b_n\}$  against  $\{c_n\}$  any most powerful level  $\alpha$  test is of the form

$$\begin{aligned} \varphi(n) &= 1 & \text{if } n < n_0 \\ \varphi(n) &= 0 & \text{if } n > n_0; \end{aligned}$$

however no uniformly least favorable distribution exists. Further  $\{a_n\}$  is least favorable for a closed set of  $\alpha$ 's which is not the union of a finite number of closed intervals.

The example is a fairly natural one to construct from the point of view of hypothesis testing. What it says about the  $\mathcal{L}^1$  norm as a byproduct is something of a surprise (at least to the originator).

**4. Acknowledgments.** Professor James G. Wendel offered many helpful suggestions and much encouragement during the development of this paper. The referee offered many suggestions which added clarity and conciseness to the exposition.

#### REFERENCES

- [1] D. A. S. FRASER, *Nonparametric Methods in Statistics*, John Wiley and Sons, New York, 1957.
- [2] PAUL R. HALMOS, *Measure Theory*, D. Van Nostrand, New York, 1950.
- [3] E. L. LEHMANN, *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.
- [4] E. L. LEHMANN, "On the existence of least favorable distributions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 408-416.
- [5] E. L. LEHMANN AND C. STEIN, "Most powerful tests of composite hypotheses I. Normal distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 495-516.
- [6] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses I. Unbiased critical regions of type  $A$  and type  $A_1$ ," *Statistical Research Memoirs*, Vol. 1 (1936), pp. 1-37.
- [7] ABRAHAM WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

# ASYMPTOTIC EFFICIENCY IN POLYNOMIAL ESTIMATION<sup>1</sup>

BY PAUL G. HOEL

*University of California, Los Angeles*

**1. Summary.** Asymptotic formulas are obtained for the generalized variance of the least squares estimates in polynomial regression under the assumption that the basic random variables are those of a stationary stochastic process, or a slight generalization of such a process. These formulas are used to study the information obtained by increasing the number of observational points in an interval and by increasing the length of the interval.

**2. Introduction.** In an earlier paper [1], some limited results were obtained on the increased efficiency of estimation in polynomial regression due to increasing the number of observational points, under the assumption that the basic random variables are correlated. These results were for two special stochastic processes only. In this paper, somewhat more general stochastic processes are studied and corresponding asymptotic formulas are obtained.

The same notation will be used here as in [1]. Thus,  $y_1, y_2, \dots, y_n$  will denote random variables associated with the fixed values  $x_1, x_2, \dots, x_n$ , and the regression polynomial will be denoted by

$$E(y_i) = \beta_0 + \beta_1 x_i + \dots + \beta_k x_i^k.$$

For convenience the interval  $(0, a)$  will be chosen as the interval over which observations are to be taken. Furthermore, in the development of the theory, the observation points  $x_1, x_2, \dots, x_n$  will be chosen to be the  $n$  equally spaced points given by the formula  $x_i = i\delta$ , where  $\delta = a/n$ .

The variables  $y_1, y_2, \dots, y_n$  will be assumed to be those of a stationary stochastic process. Thus, the  $y$ 's possess a common variance, and the correlation between  $y_i$  and  $y_j$  is a function of  $|i - j|\delta$  only. As a result, the covariance matrix  $S$  can be written in the form

$$S = \sigma^2 \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{n-1} \\ r_1 & 1 & r_1 & \cdots & r_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n-1} & r_{n-2} & r_{n-3} & \cdots & 1 \end{bmatrix}.$$

Here  $r_j$  denotes the correlation coefficient for two variables whose  $x$  values are  $j\delta$  units apart.

As before, it is necessary to introduce the spacing matrix  $X$  given by the formula

---

Received October 22, 1960; revised March 10, 1961.

<sup>1</sup> This research was supported by means of an Air Force Research Grant.

$$X = \begin{bmatrix} 1 & \delta & \delta^2 & \cdots & \delta^k \\ 1 & 2\delta & (2\delta)^2 & \cdots & (2\delta)^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & n\delta & (n\delta)^2 & \cdots & (n\delta)^k \end{bmatrix}.$$

The measure of efficiency in the estimation of the  $\beta$ 's that will be used here is the generalized variance, or, equivalently, the square of the volume of the ellipsoid of concentration. The generalized variance for best unbiased linear estimates is expressible by means of a well known formula [1] in terms of the matrices  $X$  and  $S$ .

**3. Least squares estimates.** From a theorem in the book of Grenander and Rosenblatt [2], it follows that the least squares estimates of the coefficients in polynomial regression are asymptotically efficient for stationary processes. Therefore, in studying asymptotic efficiency, least squares estimates may be used in place of Markoff estimates, provided one is dealing with stationary processes. Since least squares estimates possess a simpler generalized variance formula than Markoff estimates, it is convenient to work with them in studying the asymptotic efficiency of various spacing designs. For least squares estimates, the generalized variance of polynomial regression coefficients is given [2] by the formula

$$(1) \quad \text{G.V.} = \frac{|X'SX|}{|X'X|^2}.$$

Now consider any continuous correlation function  $\rho(t)$  defined over the closed interval  $[0, a]$ . Since it may be approximated arbitrarily closely by a finite series of the form

$$(2) \quad \rho(t) = \sum_{m=1}^N c_m \exp \{-\alpha_m t\},$$

where  $\alpha_m > 0$ , it will be assumed that the correlation function is of this type. Because  $\rho(0) = 1$ , it is necessary that  $\sum c_m = 1$ . In terms of this correlation function the value of  $r_j$  will be given by

$$r_j = \rho(j\delta) = \sum_{m=1}^N c_m \exp \{-\alpha_m j\delta\}.$$

Then  $S$  assumes the form  $S = \sigma^2(w_{ij})$  where

$$w_{ij} = \sum_{m=1}^N c_m \exp \{-\alpha_m \delta |i - j|\}.$$

Let  $S_m = (w_{ij}^{(m)})$  where  $w_{ij}^{(m)} = \exp \{-\alpha_m \delta |i - j|\}$ . Then  $S$  may be written as

$$S = \sigma^2 \sum_{m=1}^N c_m S_m.$$

As a result

$$(3) \quad X'SX = \sigma^2 \sum_{m=1}^N c_m X'S_m X.$$

Consider the typical term,  $a_{i+1,j+1}^{(m)}$ , in  $X'S_mX$ . Postmultiplying  $S_m$  by  $X$  and premultiplying the result by  $X'$  will show that  $a_{i+1,j+1}^{(m)}$  can be expressed in the form

$$a_{i+1,j+1}^{(m)} = \sum_{x=1}^n \sum_{y=1}^n (x\delta)^i (y\delta)^j e^{-\alpha_m |x\delta - y\delta|}.$$

If the substitutions  $u = x\delta$  and  $v = y\delta$  are made, it will be seen that this sum possesses the same asymptotic value with respect to  $1/\delta$  as the integral

$$(4) \quad \left(\frac{1}{\delta}\right)^2 \int_0^a \int_0^a u^i v^j e^{-\alpha_m |u-v|} du dv.$$

The typical term in  $X'SX$  will be denoted by  $b_{i+1,j+1}$ . From (3) it follows that

$$(5) \quad b_{i+1,j+1} = \sigma^2 \sum_{m=1}^N c_m a_{i+1,j+1}^{(m)},$$

and therefore that the asymptotic value of  $b_{i+1,j+1}$  is obtained by substituting the asymptotic value of  $a_{i+1,j+1}^{(m)}$  into (5). In view of formulas (2) and (4), this asymptotic value may be written in the form

$$(6) \quad b_{i+1,j+1} \sim \left(\frac{1}{\delta}\right)^2 \sigma^2 \int_0^a \int_0^a u^i v^j \rho(u-v) du dv.$$

Similar considerations will show that the typical term in  $X'X$  is given by the single sum  $\sum_{x=1}^n (x\delta)^i (x\delta)^j$ , which possesses the asymptotic value

$$(7) \quad \left(\frac{1}{\delta}\right) \int_0^a u^{i+j} du = \frac{a^{i+j+1}}{\delta(i+j+1)}.$$

It now follows from (1), (6), and (7) that the asymptotic value of the generalized variance is given by

$$\text{G. V.} \sim \frac{\left(\frac{\sigma}{\delta}\right)^2 \int_0^a \int_0^a u^i v^j \rho(u-v) du dv}{\left[\frac{1}{\delta} \frac{a^{i+j+1}}{i+j+1}\right]^2}.$$

Since the factors in  $1/\delta$  cancel, this asymptotic expression reduces to

$$(8) \quad \text{G. V.} \sim \frac{\sigma^2 \int_0^a \int_0^a u^i v^j \rho(u-v) du dv}{\left[\frac{a^{i+j+1}}{i+j+1}\right]^2}.$$

**4. Nonconstant variance.** The preceding results were based on the assumption that the process is stationary. This assumption is certainly a realistic one for many applications, at least as far as the correlation function is concerned. A more general situation, in which the variance is assumed to be a continuous function of time in the closed interval  $[0, a]$ , can be treated by the same methods as those

just employed and will be found to yield similar conclusions with respect to asymptotic efficiency.

The demonstration of this last statement can be carried out by considering  $\sigma_t$  as a polynomial in  $t$ . The resulting change in the matrix  $S_m$  will change the integral (4) to the integral

$$(9) \quad \left(\frac{1}{\delta}\right)^2 \int_0^a \int_0^a u^i v^j e^{-\alpha_m |u-v|} \sigma_u \sigma_v du dv,$$

where, say,

$$\sigma_u = \gamma_0 + \gamma_1 u + \cdots + \gamma_s u^s.$$

The substitution of this quantity in (9) will show that (9) reduces to

$$(10) \quad \left(\frac{1}{\delta}\right)^2 \sum_{p=0}^s \sum_{q=0}^s \gamma_p \gamma_q \int_0^a \int_0^a u^{i+p} v^{j+q} e^{-\alpha_m |u-v|} du dv.$$

As a result, the asymptotic value of (10) is of the same form as for (4), and therefore one would expect the same type of asymptotic efficiency properties to hold. Such properties will be considered in the next section.

**5. Adding observations.** The result given by (8) shows that when a large number of observations has been made in an interval the amount of information gained by taking additional observations is negligible. Thus, if the number of equally spaced points in an interval is doubled, which means that  $\delta$  must be replaced by  $\delta/2$ , the same asymptotic value of the generalized variance is obtained because (8) does not depend upon  $\delta$ . This holds regardless of the nature of the correlation function, provided that it is continuous. It holds not only for any stationary process but also, in view of formula (10), for a stationary process that has been modified to permit the variance to be any continuous function of  $t$ .

If the  $y$ 's are independent random variables, the generalized variance will approach zero as  $\delta$  approaches zero, and therefore it certainly pays to add observations in this case. In view of this fact, it is clear that the size of the sample needed before one can conclude that it is hardly worth while taking additional observations depends rather heavily upon the nature of the correlation function. For the purpose of observing how the value of the generalized variance changes as the nature of the correlation function changes, consider the special correlation function  $\rho(t) = e^{-\alpha t}$  that was considered in [1], and assume that  $\sigma = 1$ . Suppose the value of  $\alpha$  is changed to the value  $2\alpha$ . This is equivalent to squaring the value of the correlation coefficient between any pairs of  $y$  values, and hence in weakening the correlation relationship to this degree. For any particular value of  $\alpha$  and  $\alpha$  this effect of doubling  $\alpha$  can be determined numerically by means of formula (8); however, it is difficult to make such a comparison for a general  $\alpha$  unless  $\alpha$  is very large. Therefore, consider next an approximation to (8) which is valid for large values of  $\alpha$ .

Let

$$I_{ij} = \int_0^a \int_0^u u^i v^j e^{-\alpha(u-v)} dv du.$$

The value of  $b_{i+1,j+1}$  in (6) will then be given by the quantity  $(I_{ij} + I_{ji})/\delta^2$ . Repeated integration by parts in the first integration will show that  $I_{ij}$  can be expressed in the form

$$I_{ij} = \frac{1}{\alpha} \left\{ \frac{a^{i+j+1}}{i+j+1} - \frac{j}{\alpha} \frac{a^{i+j}}{i+j} + \frac{j(j-1)}{\alpha^2} \frac{a^{i+j-1}}{i+j-1} - \dots (-1)^j \frac{j(j-1)\dots 1}{\alpha^j} \frac{a^{i+1}}{i+1} \right\} + (-1)^{j+1} \frac{j(j-1)\dots 1}{\alpha^{j+1}} \int_0^a u^i e^{-\alpha u} du.$$

For large  $\alpha$  the first term will dominate this expression; therefore  $I_{ij}$  may be approximated by

$$I_{ij} \approx \frac{1}{\alpha} \frac{a^{i+j+1}}{i+j+1}.$$

The double integral in (8) will therefore be approximated by twice this value; consequently, for large  $\alpha$ , the generalized variance in (8) may be approximated by

$$(11) \quad G.V. \approx \frac{\left| \frac{2}{\alpha} \frac{a^{i+j+1}}{i+j+1} \right|}{\left| \frac{a^{i+j+1}}{i+j+1} \right|^{\frac{1}{2}}} = \frac{(2/\alpha)^{k+1}}{A \alpha^{(k+1)/2}},$$

where  $A = |1/(i+j+1)|$ .

In making comparisons by means of the generalized variance, it is convenient to consider the quantity introduced in [1], namely,

$$\left[ \frac{G.V.(\alpha, a)}{G.V.(2\alpha, a)} \right]^{1/(k+1)}.$$

The value of this quantity gives the number of replications of an experiment in the given interval needed to yield the same estimation information, as measured by the generalized variance, as that obtained through doubling the value of  $\alpha$ . From (11) it follows that the value of this quantity is 2 here; therefore when a large number of observations has been made, doubling the value of a large  $\alpha$  yields the same estimation information as repeating the experiment. If the typical element in the covariance matrix  $(X'X)^{-1}(X'SX)(X'X)^{-1}$  is computed, using the same approximation as before, it will be seen that doubling the value of  $\alpha$  multiplies all elements of this matrix by  $\frac{1}{2}$ ; therefore in the sense that the variance of each estimate is multiplied by  $\frac{1}{2}$ , the efficiency of estimation is doubled through doubling  $\alpha$ .

The preceding results show that doubling the number of observation points in an interval gives rise to two counteracting effects. The favorable effect arises



from doubling the size of the experiment. The unfavorable effect arises from increasing the correlation between neighboring variables. For large samples, these two effects approximately nullify each other. Since the preceding result, that adding points does not help much here, holds for a large value of  $\alpha$ , and hence for a weak correlation relationship, the advantage of adding observations in a fixed interval would be expected to be even less when there exists a strong correlation relationship. Some numerical results in this connection may be found in [1].

**6. Extending the interval.** A second form of comparison which is of interest in regression problems is that arising when the interval over which observations are to be taken is extended. This comparison for the same correlation function as in Section 5 can be made by replacing  $a$  by  $2a$  in (8) and then calculating the quantity

$$(12) \quad \left[ \frac{G. V. (\alpha, a)}{G. V. (\alpha, 2a)} \right]^{1/(k+1)}.$$

When  $\alpha$  is large, the approximation given in (11) may be used, in which case the value of this quantity will reduce to  $2^{k+1}$ . Thus, when  $\alpha$  is large and a large number of observations has been made, doubling the number of equally spaced observations by doubling the length of the interval gives approximately as much estimation information as  $2^{k+1}$  replications of the experiment in the original interval. This result was obtained in [1] by using other methods.

When  $\alpha$  is not sufficiently large to justify the use of approximation (11), numerical methods are needed to observe what effect doubling the length of the interval has on the generalized variance. Thus, calculations for  $\alpha = 1$ ,  $a = 1$ , and  $k = 2$  by means of formula (8) yielded the value 15 for the quantity given in (12). Under independence the value would have been 8; therefore there appears to be even greater advantage in extending the interval when strong correlation exists than under independence.

#### REFERENCES

- [1] PAUL G. HOEL, "Efficiency problems in polynomial estimation," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1134-1145.
- [2] ULF GRENANDER AND MURRAY ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1956.

# MAXIMUM LIKELIHOOD ESTIMATION OF A LINEAR FUNCTIONAL RELATIONSHIP

BY C. VILLEGAS

*Instituto de Matemática y Estadística, Universidad de la  
República, Montevideo, Uruguay*

**1. Introduction and summary.** We shall consider the problem of estimating a linear functional relationship

$$(1.1) \quad \alpha + \beta_1 \tau^1 + \cdots + \beta_p \tau^p = 0$$

among  $p$  variables  $\tau^1, \dots, \tau^p$  when the observed values do not satisfy it because all of them are subject to errors or fluctuations (superscripts will, in general, be indexing symbols, not powers, in this paper). Geometrically, the problem is equivalent to fitting straight lines or planes to a series of  $q$  observed points when all the coordinates are subject to error. This problem has a long history. R. J. Adcock, in two papers written in 1877 [2] and 1878 [3], solves it by minimizing the sum of squares of the orthogonal distances from the points to the hyperplane (1.1). Adcock and many other authors used the model

$$(1.2) \quad y_i = \tau_i + \epsilon_i \quad (i = 1, \dots, q),$$

where  $y_i$  and  $\tau_i$  are column vectors representing the observed and true points, and the errors  $\epsilon_i$  are independent random vectors with mean value zero. Since the  $\tau_i$  are points lying on the hyperplane (1.1), we have in matrix notation

$$(1.3) \quad \alpha + \beta \tau_i = 0 \quad (i = 1, \dots, q)$$

where  $\beta$  is a row vector with components  $\beta_1, \dots, \beta_p$ . If we assume that the  $\tau_i$  are independently drawn from a probability distribution, then the estimate of  $\beta$  obtained by Adcock is not consistent. In fact, in 1937, J. Neyman [21] pointed out that if the distribution of the true vectors  $\tau_i$  and the errors  $\epsilon_i$  is normal, then the distribution of the observed vectors  $y_i$  is also normal and, being determined by moments of the first two orders, it is not sufficient to determine the parameters  $\alpha$  and  $\beta$ ; the functional relationship (1.1) is, therefore, nonidentifiable. Several methods have been proposed to overcome this difficulty, for which the reader is referred to a recent general survey of the literature by A. Madansky [18], which also contains an extensive bibliography. One approach is to assume that we know the covariance matrix of the errors up to a numerical factor. As was shown, in general, by C. F. Gauss [8], [9], in the case of independently and normally distributed observations whose variances are known up to a numerical factor, the maximum likelihood estimate is simply the weighted least-squares estimate. This estimate of the linear functional relationship was obtained as early as 1879

Received January 25, 1960; revised April 18, 1961.

by C. H. Kummell [15], for the case in which the components  $\epsilon_i^h$  of the vectors  $\epsilon_i$  are independently distributed with variances  $\kappa^h \sigma_{hh}$ , where the  $\kappa^h$  are known constants and the  $\sigma_{hh}$  are known only up to a numerical factor. Kummell found that his estimate coincides with the estimate proposed by Adcock only in the case in which all the variances are equal.

M. J. van Uven [24] considered the case in which the errors  $\epsilon_i$  are independent and have the same multivariate normal distribution with a covariance matrix  $\Sigma$  which is known only up to a numerical factor. His method is essentially the following. He considers  $\tau^1, \dots, \tau^p$  as skew coordinates in a new, "isotropic" space in which the rectilinear orthogonal coordinates are independent and have the same variance. In the new space he then uses Adcock's principle of adjustment, namely, he chooses as the estimate the hyperplane which minimizes the sum of orthogonal distances. Later, T. Koopmans [14] showed that van Uven's estimate is the maximum likelihood estimate for the case considered. If the  $\tau_i$  are assumed independently drawn from a probability distribution, the estimate of the linear functional relationship thus obtained is consistent, but the estimate of  $\Sigma$  converges in probability to  $p^{-1}\Sigma$  (see also [16]). B. M. Dent [7] solved the maximum likelihood equations in the case in which  $\Sigma$  is not known, but, as was shown later by D. V. Lindley [16], her estimates are not consistent, and should, therefore, be rejected. More recently, J. Kiefer and J. Wolfowitz [13] showed that, under certain conditions of identifiability, when the  $\tau_i$  have a probability distribution, the method of maximum likelihood, if properly applied, yields consistent estimates of both the linear functional relationship and the probability distribution of the  $\tau_i$ . However, Kiefer and Wolfowitz do not give explicit expressions for the maximum likelihood estimates.

No difficulties with respect to the identifiability of the functional relationship or with respect to the consistency of the estimates arise if we can replicate the observations. The model is now, in matrix notation,

$$(1.4) \quad y_{ij} = \tau_i + \epsilon_{ij}.$$

In general we have  $n_i$  observed points for each of the true points  $\tau_i$ , and it is assumed that not all the  $\tau_i$  lie on a translated subspace of dimension smaller than  $p - 1$ . Obviously this implies that  $q \geq p$ . This model has been considered previously by G. W. Housner and J. F. Brennan [12], J. W. Tukey [23] and by F. S. Acton [1] (see also [11] and [25]). If we assume that the errors  $\epsilon_{ij}$  are independently and normally distributed with a known covariance matrix  $\Sigma$ , we lose nothing if we consider only the averages  $y_{i\cdot} = n_i^{-1} \sum_j y_{ij}$ . We have then the same model (1.2), with the only difference that  $y_i$  is replaced by  $y_{i\cdot}$ . If, however, the covariance matrix is not known, we can now obtain in the usual way an estimate  $S$  of  $\Sigma$ . F. S. Acton [1] suggested the use of  $S$  instead of  $\Sigma$  in the estimate obtained by the method of maximum likelihood in the case of known  $\Sigma$ . In this paper it will be shown that the estimate thus obtained is the maximum likelihood estimate when  $\Sigma$  is unknown.

If the design is a (in general incomplete) block design, we have, if the treatment  $i$  is applied on the block  $j$ ,

$$(1.5) \quad y_{ij} = \tau_i + b_j + \epsilon_{ij},$$

where  $b_j$  is a column vector representing the block effect. Considering the block effects  $b_j$  as unknown constants, we get in the usual way the intrablock estimates  $t_i$  of the treatment effects  $\tau_i$ . Then, the same equation (1.2) still holds if  $y_i$  is replaced by  $t_i$ , but in general the errors  $\epsilon_i$  and consequently also the estimates  $t_i$  will no longer be independent. If the design consists of  $r$  replications of a basic design, then the covariance of two errors  $\epsilon_i, \epsilon_{i'}$  will be given by

$$(1.6) \quad \text{cov}(\epsilon_i, \epsilon_{i'}) = \frac{\kappa^{ii'} \Sigma}{r}$$

where  $\kappa^{ii'}$  are known coefficients and the matrix  $\Sigma$  is unknown.

In this paper maximum likelihood estimates for  $\Sigma$  and the parameters of the linear functional relationship will be found for the case in which (1.6) holds. It will be shown that the maximum likelihood estimates  $\hat{\alpha}, \hat{\beta}$  in the case of unknown  $\Sigma$  are obtained from the corresponding estimates in the case of known  $\Sigma$  by simply replacing  $\Sigma$  by the linear regression estimate  $S$ . In the last Section it will be shown that if the maximum likelihood method is applied directly to the variables  $y_{ij}$  instead of the variables  $t_i$  and  $S$ , then the same estimate  $\hat{\beta}$  is obtained, but the estimate of  $\Sigma$  is multiplied by  $1 - k^{-1} + N^{-1}$ , where  $k$  is the number of experimental units in each block and  $N$  is the total number of experimental units. All of the estimates obtained are consistent, with the exception only of the estimate of  $\Sigma$  obtained by the direct approach in the last Section, which converges to  $(1 - k^{-1})\Sigma$ .

**2. The model.** We shall consider now in more detail the intrablock analysis of a (in general incomplete) block design to which the additive model (1.5) applies. We shall assume that errors coming from different experimental units are independent, and that the errors coming from a single experimental unit have a multivariate normal distribution with zero means and covariance matrix  $\Sigma = \{\sigma_{hh'}\}$ . Therefore,

$$(2.1) \quad \text{cov}(\epsilon_{ij}^h, \epsilon_{ij}^{h'}) = \sigma_{hh'},$$

where  $\epsilon_{ij}^h (h = 1, \dots, p)$  are the components of  $\epsilon_{ij}$ . If we do not consider the linear functional relationship (1.1), the estimation of  $\tau_i$  and  $\Sigma$  is simply a linear regression problem. In order to arrive at a unique solution  $t_1, \dots, t_q$ , it is usual to add some arbitrary linear restriction, say,

$$(2.2) \quad \sum_i \omega_i t_i = 0, \quad \sum_i \omega_i \neq 0.$$

where  $\kappa^{ii}$  are known coefficients and the matrix  $\Sigma$  is unknown.

It is known ([5] Section 8.2) that the linear regression estimates  $t_i$  are linear combinations of the observed vectors  $y_{ij}$ . If the design consists of  $r$  replications of a basic design, then the covariance of two estimates  $t_i, t_{i'}$  is given by (1.6),

where the  $\kappa^{ii'}$  are known coefficients that depend on the basic design and the linear restriction (2.2). Since the errors are assumed to be normally distributed, it follows that the  $t_i$  have a multivariate distribution with means  $\tau_i$  and covariances given by (1.6). Finally, the linear regression estimate of  $\sigma_{hh'}$  is

$$(2.3) \quad s_{hh'} = \frac{1}{\nu} \sum'_{ij} e_{ij}^h e_{ij}^{h'}$$

where  $\nu = N - q - b + 1$ ,  $b$  is the number of blocks,  $e_{ij}^h$  is the linear regression estimate of the error  $\epsilon_{ij}^h$ , and the prime over the summation sign indicates that the sum must be extended over all pairs  $(i, j)$  such that treatment  $i$  appears on block  $j$ . It is known ([5], Section 8.2) that the estimated covariance matrix  $S = \{s_{hh'}\}$  is independent of the  $t_i$  and has a Wishart distribution with mean value  $\Sigma$  and  $\nu$  degrees of freedom.

We have then the following linear functional relationship model. The  $p$ -dimensional random variables  $t_1, \dots, t_q$  have a multivariate normal distribution, with means  $\tau_1, \dots, \tau_q$  that satisfy (1.3) and covariances

$$(2.4) \quad \text{cov}(t_i, t_{i'}) = \frac{\kappa^{ii'}}{r} \Sigma$$

where  $r$  and the  $\kappa^{ii'}$  are known coefficients, and  $\Sigma$  is unknown. The matrix  $S$  is an unbiased estimate of  $\Sigma$ , is independent of the  $t_i$  and has a Wishart distribution with a number of degrees of freedom  $\nu$  which tends to infinity when  $r \rightarrow \infty$ ; the quotient  $r/\nu$  converging to a positive limit.

The matrix  $K = \{\kappa^{ii'}\}$  is always nonnegative because, for a given  $h$ ,  $r^{-1}\sigma_{hh}K$  is the covariance matrix of the  $h$ th components  $t_1^h, \dots, t_q^h$  of  $t_1, \dots, t_q$ . If the  $t_i$  are not subject to any linear restriction like (2.2), then for any  $h$  the distribution of  $t_1^h, \dots, t_q^h$  is of rank  $q$  (see for example [6], p. 297) and the matrix  $K$  is positive definite. If there is only one linear restriction (2.2), the matrix  $K$  is of rank  $q - 1$  and

$$(2.5) \quad K\omega = 0,$$

where  $\omega$  is the column vectors the components of which are  $\omega_1, \dots, \omega_q$  (see for instance [17]).

**3. Covariance matrix known.** We shall consider in the first place the case in which the covariance matrix  $\Sigma$  is a known positive definite matrix and the  $t_i$  are not subject to any linear restriction (and consequently  $K$  is a positive definite matrix).

From (2.4) it follows that the covariance matrix of all the variables  $t_i^h$  is  $r^{-1}K \otimes \Sigma$ , where the symbol  $\otimes$  denotes the Kronecker product of two matrices. The determinant of this covariance matrix is  $r^{-pq}|K|^p|\Sigma|^q$ . Therefore the probability density function for  $t_1^h, \dots, t_q^h$  is, up to a numerical factor, equal to

$$(3.1) \quad |K|^{-1p} |\Sigma|^{-1q} e^{-1/2r},$$

where, if  $K^{-1} = \{\kappa_{ii'}\}$ ,

$$Q = \sum_{i,i'} \kappa_{ii'} (t_i - \tau_i)' \Sigma^{-1} (t_{i'} - \tau_{i'}).$$

We shall denote the trace of a matrix  $X$  by  $\text{tr } X$ . Then, in matrix notation,

$$(3.2) \quad Q = \text{tr } \Sigma^{-1} (t - \tau) K^{-1} (t - \tau)',$$

where  $t$  is the  $p \times q$  matrix the  $i$ th column of which is  $t_i$ , and similarly,  $\tau$  is the  $p \times q$  matrix the  $i$ th column of which is  $\tau_i$ . The maximum likelihood estimates of  $\alpha$ ,  $\beta$  and  $\tau$  are the values  $\hat{\alpha}$ ,  $\hat{\beta}$  and  $\hat{\tau}$  that minimize (3.2) subject to the conditions (1.3) which may be written

$$(3.3) \quad \alpha u + \beta \tau = 0,$$

where  $u$  is the row vector the  $q$  components of which are equal to 1. This problem was solved by Koopmans [14] for the particular case  $K = I$ . In what follows, unless otherwise specified, the symbols  $\alpha$ ,  $\beta$  and  $\tau$  will denote mathematical variables (not true values or true parameters). In the first place we shall find the minimum of (3.2) subject to the condition (3.3), for given values of  $\alpha$  and  $\beta$ . Since  $\Sigma$  and  $K$  are positive definite, uniquely defined positive square roots  $\Sigma^{1/2}$  and  $K^{1/2}$  exist (see [10], p. 166). Consider the change of variable

$$(3.4) \quad \delta = \Sigma^{-1/2} (t - \tau) K^{-1/2}.$$

Since  $\text{tr } XY = \text{tr } YX$ , (3.2) may be written, if  $\delta_i$  is the  $i$ th column of  $\delta$ ,

$$(3.5) \quad Q = \text{tr } \delta \delta' = \sum_i \delta_i' \delta_i,$$

that is,  $Q$  is the sum of the squares of the distances from the origin to the points  $\delta_i$ . The condition (3.3) is, in the new variables,

$$(3.6) \quad \gamma_i - \beta \Sigma^{-1/2} \delta_i = 0, \quad (i = 1, \dots, q)$$

where  $\gamma_i$  is the  $i$ th component of  $(\alpha u + \beta t) K^{-1/2}$ . We have to minimize the sum of squares of the distances from the origin to the points  $\delta_i$ , subject to the condition that each  $\delta_i$  lies on the corresponding hyperplane (3.6). Note that these are, in general,  $q$  parallel hyperplanes. This was the principle from which M. J. van Uven [24] derived his estimates for the case  $K = I$ . Obviously, the minimum is reached when  $\delta_i$  is on the perpendicular from the origin to the hyperplane (3.6) and, therefore,

$$\delta_i = \gamma_i \frac{\Sigma^{1/2} \beta'}{\beta \Sigma \beta'}.$$

Going back to the old variables, we have

$$(3.7) \quad t - \hat{\tau} = \frac{\Sigma \hat{\beta}'}{\hat{\beta} \Sigma \hat{\beta}'} (\hat{\alpha} u + \hat{\beta} t).$$



By substitution of (3.7) into (3.2) it follows that the minimum value of  $Q$ , for given values of  $\alpha$  and  $\beta$ , is

$$(3.8) \quad Q_1 = \frac{(\alpha u + \beta t)K^{-1}(\alpha u + \beta t)'}{\beta \Sigma \beta'}$$

We shall now find the minimum value of  $Q_1$  for a given  $\beta$ . By differentiation we have  $(\alpha u + \beta t)K^{-1}u' = 0$ . If we set

$$w = \frac{K^{-1}u'}{uK^{-1}u'}$$

and  $t = tw = \sum_i t_i w_i$ , where  $w_1, \dots, w_q$  are the components of  $w$ , we get

$$(3.9) \quad \hat{\alpha} = -\hat{\beta}t.$$

Since  $Q_1$  tends to  $+\infty$  when  $\alpha \rightarrow \pm\infty$ , it follows that the value  $\alpha$  given by (3.9) minimizes  $Q_1$ . From the definition of  $w$  we have  $uw = \sum w_i = 1$ , so that if, as happens with the usual matrices,  $K$ , the  $w_i$  are nonnegative, then  $t$  is a weighted average of the vectors  $t_i$ , with weights,  $w_i$ . If we write  $\Delta t_i = t_i - t$  and  $\Delta t = t - tu$ , we have at the minimum  $\alpha u + \beta t = \beta \Delta t$ , and, therefore, the minimum value of  $Q_1$  is

$$(3.10) \quad Q_2 = \frac{\beta F \beta'}{\beta \Sigma \beta'}$$

where

$$(3.11) \quad F = \Delta t K^{-1} (\Delta t)'$$

is a nonnegative matrix.

We now have to find the vector  $\hat{\beta}$  which minimizes  $Q_2$ . Consider the change of variable  $\beta = \mu \Sigma^{-1}$ . Substituting into (3.10)

$$(3.12) \quad Q_2 = \frac{\mu \Sigma^{-1} F \Sigma^{-1} \mu'}{\mu \mu'}$$

The vector  $\hat{\mu}$  which minimizes this expression is obviously any proper vector of the smallest proper value of the nonnegative matrix  $\Sigma^{-1} F \Sigma^{-1}$ ; that is,  $\hat{\mu}$  is given by

$$\hat{\mu}(\Sigma^{-1} F \Sigma^{-1} - \lambda I) = 0,$$

where  $\lambda$  is the smallest root of the equation,

$$|\Sigma^{-1} F \Sigma^{-1} - \lambda I| = 0.$$

In computations the following equivalent equations may be preferred. The maximum likelihood estimate of  $\beta$  is given by the equation

$$(3.13) \quad \hat{\beta}(F - \lambda \Sigma) = 0,$$

where  $\lambda$  is the smallest root of

$$(3.14) \quad |F - \lambda \Sigma| = 0.$$

**4. Covariance matrix unknown.** We assume now that the covariance matrix  $\Sigma$  is unknown, but that we have an estimate  $S$  which is independent of  $t$  and has a Wishart distribution with mean value  $\Sigma$  and  $\nu$  degrees of freedom, the probability density of which is proportional to

$$(4.1) \quad |\Sigma|^{-\frac{1}{2}\nu} |S|^{\frac{1}{2}(\nu-p-1)} \exp -\frac{1}{2}\nu \operatorname{tr} \Sigma^{-1} S$$

for  $S$  positive definite and 0 otherwise, where  $S = \{s_{hh'}\}$ . It is assumed, as before, that  $\Sigma$  is positive definite, and that the  $t_i$  are not subject to any linear restriction (and, therefore,  $K$  is positive definite). We shall consider only the case, which happens with probability 1, in which  $S$  is positive definite.

The joint probability density of all of the variables  $t_i^A$ ,  $s_{hh'}$  is proportional to the product of (3.1) and (4.1). The maximum likelihood estimates are the values  $\hat{\tau}$ ,  $\hat{\alpha}$ ,  $\hat{\beta}$  and  $\hat{\Sigma}$  that maximize

$$(4.2) \quad |\Sigma|^{-\frac{1}{2}(q+\nu)} \exp -\frac{1}{2} \operatorname{tr} \Sigma^{-1} [\nu S + r(t - \tau) K^{-1} (t - \tau)']$$

subject to the condition (3.3). Instead of maximizing (4.2) we can minimize

$$(4.3) \quad \operatorname{tr} \Sigma^{-1} [\nu S + r(t - \tau) K^{-1} (t - \tau)'] - (q + \nu) \log |\Sigma^{-1}|.$$

We may in the first place keep  $\alpha$ ,  $\beta$  and  $\tau$  fixed and find the value of  $\Sigma$  which minimizes (4.3). By the Lemma 3.2.2 of Anderson's book [4] the maximum likelihood estimate of  $\Sigma$  is

$$(4.4) \quad \hat{\Sigma} = (q + \nu)^{-1} [\nu S + r(t - \tau) K^{-1} (t - \tau)']$$

where  $\hat{\tau}$  is the maximum likelihood estimate of  $\tau$  (based on maximum likelihood estimates of  $\alpha$  and  $\beta$  and the restraint (3.3)). Substituting this estimate of  $\Sigma$  (as a function of  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\tau}$ ) into (4.3) we see that  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\tau}$  must minimize

$$(4.5) \quad |\nu S + r(t - \tau) K^{-1} (t - \tau)'|$$

subject to the aforementioned restraint. Consider the change of variable

$$(4.6) \quad \delta = S^{-1} (t - \tau) K^{-1}$$

We have then to minimize

$$(4.7) \quad |\nu I + r \delta \delta'|$$

subject to the following conditions, similar to (3.5),

$$(4.8) \quad \gamma_i - \beta S^A \delta_i = 0 \quad (i = 1, \dots, q),$$

where, as in Section 3,  $\gamma_i$  is the  $i$ th component of  $(\alpha u + \beta t) K^{-1}$  and  $\delta_i$  is the  $i$ th column of  $\delta$ . We shall find in the first place the minimum of (4.7) for fixed values of  $\alpha$  and  $\beta$ . The expression (4.7) may be written

$$\nu^p + D_1 \nu^{p-1} r + D_2 \nu^{p-2} r^2 + \dots + D_p r^p$$

where  $D_h$  is the sum of all the principal minors of order  $h$  of the matrix  $\delta\delta'$ . The elements of the matrix  $\delta\delta'$  are the product-moments with respect to the origin of the system of points  $\delta_i$ , and consequently all the principal minors of  $\delta\delta'$ , and a fortiori all the coefficients  $D_h$ , are non-negative. In particular,

$$(4.9) \quad D_1 = \text{tr } \delta\delta' = \sum_i \delta_i' \delta_i,$$

is the sum of the squares of the distances from the origin to the points  $\delta_i$ . This is a minimum when all the points  $\delta_i$  are on the perpendicular from the origin to the hyperplanes (4.8). Since these points are on a straight line which goes through the origin, it is easily seen that at these points all of the principal minors of  $\delta\delta'$  of order  $\geq 2$ , and consequently, also  $D_2, \dots, D_p$ , vanish simultaneously. Therefore, minimizing (4.7) is equivalent to minimizing (4.9) subject to the conditions (4.8). The same problem was solved in Section 3, with the only difference that we now have  $S$  instead of  $\Sigma$ . Therefore, the maximum likelihood estimate of  $\beta$  is given by the equation

$$(4.10) \quad \hat{\beta}(F - lS) = 0$$

where  $l$  is the smallest root of

$$(4.11) \quad |F - lS| = 0.$$

Equivalently,  $\hat{\beta} = \hat{m}S^{-1}$ , where  $\hat{m}$  is any proper vector of the minimum proper value  $l$  of the nonnegative matrix  $S^{-1}FS^{-1}$ . The maximum likelihood estimate of  $\alpha$  is given by the same equation (3.9) as before, the maximum likelihood estimate of  $\tau$  is

$$(4.12) \quad \hat{\tau} = t - \frac{S\hat{\beta}'\hat{\beta}\Delta t}{\hat{\beta}S\hat{\beta}'}$$

and the maximum likelihood estimate of  $\Sigma$  is given by (4.4), or, equivalently, by

$$(4.13) \quad \hat{\Sigma} = \frac{1}{q + r} \left( vS + r \frac{S\hat{\beta}'\hat{\beta}S}{\hat{\beta}S\hat{\beta}'} \right).$$

**5. Consistency.** When the number  $r$  of replications tends to infinity,  $S$  converges in probability to  $\Sigma$  and  $F$  converges in probability to  $\Phi = \Delta\tau K^{-1}(\Delta\tau)'$  where  $\Delta\tau = \tau - \tau.u$  and  $\tau_i = \tau.w$ . The direction of the true vector  $\beta$  is the only direction orthogonal to all the vectors  $\Delta\tau_i = \tau_i - \tau$ , because all points  $\tau_i$  lie on the hyperplane (1.1) but do not lie on any translated subspace of smaller dimension. Consequently, up to an arbitrary factor, the true vector  $\beta$  is the only vector such that  $\beta\Delta\tau = 0$ , and, since  $K^{-1}$  is a positive definite matrix, it is also the only vector such that  $\beta\Phi\beta' = 0$ . But, since  $\Phi$  is a nonnegative matrix,  $\beta$  is the only vector (up to an arbitrary factor) such that  $\beta\Phi = 0$ . Let  $\mu$  be defined by  $\beta = \mu\Sigma^{-1}$ . Then  $\mu$  is the only vector (up to an arbitrary factor) such that  $\mu\Sigma^{-1}\Phi\Sigma^{-1} = 0$ . Therefore, the matrix  $\Sigma^{-1}\Phi\Sigma^{-1}$  is singular, the smallest proper value is 0, and its only proper vector is  $\mu$ . Since the proper vector is a continuous

function of the matrix,  $\hat{m}$  converges in probability<sup>1</sup> to  $\mu$  and consequently  $\hat{\beta}$  converges in probability to  $\beta$  (up to an arbitrary factor). Since  $\beta \Delta t$  converges in probability to  $\beta \Delta \tau = 0$ , it follows from (4.12) that  $\hat{\tau}$  converges in probability to the true matrix  $\tau$ . It follows then easily from (3.9) that  $\hat{\alpha}$  converges to  $\alpha$  and from (4.4) that  $\hat{\Sigma}$  converges to  $\Sigma$  (since by assumption, the quotient  $r/\nu$  converges to a positive limit).

**6. Homogeneous linear functional relationship.** We assume as before that there is no linear restriction (2.2), and, therefore,  $K$  is positive definite. If it is known that  $\alpha = 0$ , the equation (3.8) may be written simply

$$(6.1) \quad Q_1 = \frac{\beta t K^{-1} t' \beta'}{\beta \Sigma \beta'}.$$

Instead of defining  $F$  by (3.11) we shall define  $F$  by

$$(6.2) \quad F = t K^{-1} t',$$

and we have

$$(6.3) \quad Q_1 = \frac{\beta F \beta'}{\beta \Sigma \beta'},$$

which is similar to (3.10). We can then proceed by the same method that was employed in Section 3. The results and formulae obtained there and in Section 4 will also apply to the homogeneous linear functional relationship case, provided that it is understood that  $F$  is given by (6.2) and not by (3.11).

**7. Linear restrictions.** We shall now assume that the  $t_i$  are subject to a known single linear restriction (2.2). In matrix notation

$$(7.1) \quad t \omega = 0.$$

In this section we shall assume that the coefficients  $\omega_i$  are normalized so that

$$(7.2) \quad u \omega = 1.$$

Since  $t$  is assumed to verify (7.1), it follows that it is impossible that  $t = u$ . It follows also that we have

$$(7.3) \quad \tau \omega = 0.$$

If we multiply (3.3) on the right by  $\omega$  we obtain, by (7.2) and (7.3),

$$(7.4) \quad \alpha = 0.$$

By (3.3) we have then

$$(7.5) \quad \beta \tau = 0.$$

<sup>1</sup> Because, if the vector valued function  $f$  is continuous at the point  $a$ , and if the random vector  $x$  converges in probability to  $a$ , then  $f(x)$  converges in probability to  $f(a)$ . This result follows as a special case ( $y_N$  constant) from Corollary 2 of [20]. (See also Lemma V of [19]).

Since there is only one linear restriction (2.2), the matrix  $K$  is of rank  $q - 1$ , as was pointed out in Section 2, and consequently, the previous theory cannot be applied directly in this case. Let  $L = \{L_{ii'}\}$  be a nonsingular  $q \times q$  matrix whose last column is the vector  $\omega$ , i.e.,

$$(7.6) \quad L_{iq} = \omega_i \quad (i = 1, \dots, q),$$

and such that all other columns have a sum equal to zero, i.e.,

$$(7.7) \quad \sum_{i'} L_{ii'} = 0 \quad (i = 1, \dots, q - 1).$$

Consider the new variables

$$(7.8) \quad t_i^{*h} = \sum_{i'} t_{i'}^h L_{ii'}.$$

By (2.2) and (7.6)

$$(7.9) \quad t_q^{*h} = 0 \quad (h = 1, \dots, p).$$

If we denote by  $t^*$  the  $p \times (q - 1)$  matrix the elements of which are the  $t_i^{*h}$  with  $i \neq q$ , and by  $\tau^*$  and  $\tau_i^{*h}$  the corresponding true values, we have in matrix notation

$$(7.10) \quad (t^* | 0) = tL, \quad (\tau^* | 0) = \tau L.$$

Therefore, if we multiply (7.5) on the right by  $L$ , we obtain

$$(7.11) \quad \beta \tau^* = 0.$$

Consequently, the new variables  $\tau_i^{*h}$  ( $i \neq q$ ) satisfy an homogeneous linear functional relationship with the same parameter  $\beta$ . It can be easily seen that the covariance matrix of the  $pq$  variables  $t_i^{*h}$  is  $r^{-1}L'KL \otimes \Sigma$ . From (2.5) and (7.6) it follows that

$$(7.12) \quad L'KL = \begin{pmatrix} K^* & 0 \\ 0 & 0 \end{pmatrix},$$

where  $K^*$  is a  $(q - 1) \times (q - 1)$  matrix. Since  $L$  is nonsingular, and  $K$  has rank  $q - 1$ , it follows that  $L'KL$  has also rank  $q - 1$  and, therefore,  $K^*$  is nonsingular. Since  $r^{-1}K^* \otimes \Sigma$  is the covariance matrix of the  $p(q - 1)$  variables  $t_i^{*h}$  ( $i \neq q$ ), in order to find the maximum likelihood estimates, in the case of unknown  $\Sigma$ , we have to minimize

$$(7.13) \quad \text{tr } \Sigma^{-1}(\nu S + r(t^* - \tau^*)K^{*-1}(t^* - \tau^*)') - (q + \nu)\log |\Sigma^{-1}|$$

subject to the only restriction (7.11). This is an homogeneous linear functional-relationship problem of the type discussed in the previous section. Therefore, the maximum likelihood estimates  $\hat{\beta}$ ,  $\hat{\Sigma}$  are given by (4.10) and (4.13), where  $l$  is, as before, the smallest root of (4.11) and  $F$  is given by

$$(7.14) \quad F = t^*K^{*-1}t^{*'}.$$

We shall now show that we also have

$$(7.15) \quad F = t(K + \kappa_0 u' u)^{-1} t',$$

where  $\kappa_0$  is an arbitrary number different from zero. From (7.12) and (7.7) it follows that

$$(7.16) \quad L'(K + \kappa_0 u' u)L = \begin{pmatrix} K^* & 0 \\ 0 & \kappa_0 \end{pmatrix},$$

and, consequently,  $K + \kappa_0 u' u$  is nonsingular. Moreover, the right-hand member of (7.15) is equal to

$$t[L'(K + \kappa_0 u' u)L]^{-1} L' t'$$

and, therefore, by (7.10) and (7.16) is equal to the right-hand member of (7.14). In practical applications the expression (7.15) will be used with preference to (7.14). Moreover, in the case of balanced designs, all of the elements that are not in the diagonal of  $K$  will have a common value  $\kappa' \neq 0$ . By choosing  $\kappa_0 = -\kappa'$ , the matrix  $K + \kappa_0 u' u$  is a diagonal matrix and, consequently, the computations are considerably simplified.

A similar argument shows that the maximum likelihood estimates are also the values that minimize

$$(7.17) \quad \text{tr } \Sigma^{-1}[\nu S + r(t - \tau)(K + \kappa_0 u' u)^{-1}(t - \tau)'] - (q + \nu) \log |\Sigma^{-1}|$$

subject to the conditions (7.3) and (7.5).

**8. Intrablock analysis: direct approach.** We shall now estimate the linear functional relationship by applying directly the maximum likelihood method to the model (1.5), considering the coefficients  $b_j$  as unknown constant vectors (intrablock analysis). In order to arrive at a unique solution  $\hat{\tau}$  we add as usual the linear restriction (2.2). It follows then, as was shown in the previous section, that  $\alpha = 0$ . As was already mentioned in Section 2, we assume that the errors coming from different experimental units are independent, and that the errors coming from a single experimental unit have a multivariate normal distribution with zero means and covariance matrix  $\Sigma$ . If there are  $N$  experimental units, then the probability density for all  $pN$  variables  $y_{ij}^h$  is proportional to

$$(8.1) \quad |\Sigma|^{-1N} \exp -\frac{1}{2} \sum_{hh'} \sigma^{hh'} Q_{hh'}$$

where

$$(8.2) \quad Q_{hh'} = \sum_{ij}^t \epsilon_{ij}^h \epsilon_{ij}^{h'}.$$

The maximum likelihood estimates are the values that maximize (8.1) subject to the conditions (1.3) and (2.2), or, equivalently, the conditions (7.3) and (7.5). Let  $\bar{y}_{.j}^h$ ,  $\bar{t}_{.j}^h$  and  $\bar{\tau}_{.j}^h$  denote the average of yields, estimated and true treatment effects for the  $h$ th variable over all experimental units of block  $j$ . Define



the adjusted yields and adjusted treatment effects by

$$\bar{y}_{ij}^h = y_{ij}^h - \bar{y}_{\cdot j}^h, \quad \bar{t}_{ij}^h = t_{ij}^h - \bar{t}_{\cdot j}^h, \quad \bar{\tau}_{ij}^h = \tau_{ij}^h - \bar{\tau}_{\cdot j}^h,$$

where  $\tau_i^h$  is the  $h$ th component of  $\tau_i$ .

It can be easily shown that, if  $\hat{\tau}_i^h$  is the maximum likelihood estimate of  $\tau_i^h$ , then the maximum likelihood estimate of  $b_j^h$  is  $\hat{b}_j^h = \bar{y}_{\cdot j}^h - \hat{\tau}_{\cdot j}^h$ , where  $\hat{\tau}_{\cdot j}^h$  denotes the average of the  $\hat{\tau}_i^h$  for all treatments occurring in block  $j$ . By substitution into (8.1) and (8.2) it follows that the maximum likelihood estimates  $\hat{\beta}$ ,  $\hat{\tau}$ ,  $\hat{\Sigma}$  are the values that maximize, subject to the conditions (7.3) and (7.5) the expression

$$(8.3) \quad |\Sigma|^{-1/2} \exp -\frac{1}{2} \sum_{h,h'} \sigma^{hh'} Q'_{hh'}$$

where

$$(8.4) \quad Q'_{hh'} = \sum_{i,j} (\bar{y}_{ij}^h - \bar{\tau}_{ij}^h)(\bar{y}_{ij}^{h'} - \bar{\tau}_{ij}^{h'}).$$

Suppose that we have numbered serially the  $N$  experimental units, and let  $\bar{y}_n^h$ ,  $\bar{\tau}_n^h$ ,  $\bar{t}_n^h$  denote the adjusted yield, and the adjusted treatment effects (true and estimated) for the  $n$ th experimental unit and the  $h$ th characteristic being measured. If  $\bar{y}^h$ ,  $\bar{\tau}^h$ ,  $\bar{t}^h$  denote the row vectors the  $N$  components of which are the corresponding experimental unit values, we have

$$(8.5) \quad Q'_{hh'} = (\bar{y}^h - \bar{\tau}^h)(\bar{y}^{h'} - \bar{\tau}^{h'})'.$$

From the definition of  $\bar{\tau}^h$  it follows that

$$(8.6) \quad \bar{\tau}^h = \sum_i c_i \tau_i^h, \quad \bar{t}^h = \sum_i c_i t_i^h,$$

where the  $c_i$  are row vectors that depend only on the experimental design. Since the  $t_i^h$  are the values that minimize  $Q'_{hh}$  subject to the only conditions

$$(8.7) \quad \sum_i \omega_i t_i^h = 0 \quad (h = 1, \dots, p)$$

by differentiation with respect to  $\tau_i^h$  we have, if  $\lambda$  is a Lagrange multiplier,

$$c_i(\bar{y}^h - \bar{t}^h)' + \lambda \omega_i = 0.$$

If we multiply this expression by  $\bar{t}_i^h$  and add for  $i = 1, \dots, q$  we have by (8.6) and (8.7)

$$(8.8) \quad \bar{t}^h(\bar{y}^h - \bar{t}^h)' = 0.$$

If we multiply instead by  $\bar{\tau}_i^h$  we have

$$(8.9) \quad \bar{\tau}^h(\bar{y}^h - \bar{t}^h)' = 0.$$

By (8.8) and (8.9) it follows then from (8.5) that

$$Q'_{hh'} = (\bar{y}^h - \bar{t}^h)(\bar{y}^{h'} - \bar{t}^{h'})' + (\bar{t}^h - \bar{\tau}^h)(\bar{t}^h - \bar{\tau}^h)'$$

and by (2.3)

$$(8.10) \quad Q'_{hh'} = \nu s_{hh'} + (\bar{t}^h - \bar{\tau}^h)(\bar{t}^{h'} - \bar{\tau}^{h'})'.$$

Let  $\lambda_{ii'}$  be the number of blocks where both treatments  $i$  and  $i'$  are applied (and consequently  $\lambda_{ii}$  will be the number of replications of the treatment  $i$ ). Let  $\Lambda = \{\lambda_{ii'}\}$  and let  $\Lambda_d$  be the diagonal matrix whose diagonal is  $\lambda_{11}, \dots, \lambda_{qq}$ . Assume that all blocks contain  $k$  experimental units and define the matrix  $\tilde{K} = \{\tilde{k}_{ii'}\}$  by

$$(8.11) \quad \tilde{K} = \omega\omega' + \frac{1}{r} \left( \Lambda_d - \frac{\Lambda}{k} \right).$$

It can be easily shown that

$$(\bar{t}^h - \bar{\tau}^h)(\bar{t}^{h'} - \bar{\tau}^{h'})' = r \sum_{i,i'} (\tilde{k}_{ii'} - \omega_i \omega_{i'}) (t_i^h - \tau_i^h)(t_{i'}^{h'} - \tau_{i'}^{h'}).$$

Therefore, in matrix notation we have, by (7.1) and (7.3),

$$(8.12) \quad \sum_{h,h'} \sigma^{hh'} Q'_{hh'} = \text{tr } \Sigma^{-1} [\nu S + r(t - \tau)\tilde{K}(t - \tau)'].$$

Let  $\tilde{Y}$  be the matrix of the adjusted total yields  $\tilde{Y}_i^h = \sum_{i'} \tilde{y}_{ii'}^h$ . It can be shown that (see for instance [4], p. 251)  $\tilde{Y} = t(\Lambda_d - \Lambda/k)$ . Then, by (8.11),

$$(8.13) \quad \tilde{Y} = r t (\tilde{K} - \omega\omega').$$

We shall show that the system of equations (7.1) and (8.13) is equivalent to

$$(8.14) \quad r t \tilde{K} = \tilde{Y}.$$

It is obvious that (8.14) is a consequence of (7.1) and (8.13). From the definitions of  $\Lambda$  and  $\tilde{Y}$  it follows that  $\Lambda u' = k \Lambda_d u'$  and  $\tilde{Y} u' = 0$ . By (8.11) we have then

$$(8.15) \quad \tilde{K} u' = \omega u \omega.$$

Therefore, if we multiply (8.14) on the right by  $u'$ , we obtain the equation (7.1). From (7.1) and (8.14) the equation (8.13) follows immediately. We assume now that the design of the experiment is such that the system of equations (7.1) and (8.13) has a unique solution,  $t$ . Then it follows that (8.14) has a unique solution, and, therefore, that  $\tilde{K}$  is nonsingular. If  $\tilde{K}^{-1} = \{\tilde{k}^{ii'}\}$  is the inverse matrix, then

$$t_i^h = \frac{1}{r} \sum_{i'} \tilde{k}^{ii'} \tilde{Y}_{i'}^h,$$

and, therefore,

$$\text{cov}(t_i^h, t_{i'}^{h'}) = \frac{1}{r^2} \sum_{j,j'} \tilde{k}^{ij} \tilde{k}^{i'j'} \text{cov}(Y_j^h, Y_{j'}^{h'}).$$

It can be shown that

$$\text{cov}(Y_j^h, Y_{j'}^{h'}) = r(\tilde{k}_{jj'} - \omega_j \omega_{j'}) \sigma_{hh'}$$

and, therefore, (1.6) holds, with

$$\kappa^{ii'} = \bar{\kappa}^{ii'} - \left( \sum_j \bar{\kappa}^{ij} \omega_j \right) \left( \sum_{j'} \bar{\kappa}^{i'j'} \omega_{j'} \right).$$

In matrix notation  $K = \bar{K}^{-1} - \bar{K}^{-1} \omega (\bar{K}^{-1} \omega)'$  and by (8.15)

$$(8.16) \quad \bar{K}^{-1} = K + \kappa_0 u' u,$$

where  $\kappa_0 = (u\omega)^{-2}$  is an arbitrary positive number. By substitution into (8.12) we have

$$\sum_{h,h'} \sigma^{hh'} Q'_{hh'} = \text{tr } \Sigma^{-1} [\nu S + r(t - \tau)(K + \kappa_0 u' u)^{-1}(t - \tau)'].$$

To maximize (8.3) subject to the conditions (7.3) and (7.5) is then equivalent to minimize

$$(8.17) \quad \text{tr } \Sigma^{-1} [\nu S + r(t - \tau)(K + \kappa_0 u' u)^{-1}(t - \tau)'] - N \log |\Sigma^{-1}|$$

subject to the same conditions. But in Section 7 the same problem was solved, with the only difference being that we now have  $N$  instead of  $q + \nu = N - b + 1$ , where  $b$  is the number of blocks. The estimate  $\hat{\beta}$  is, therefore, the same as before, but instead of (4.13) we have now

$$(8.18) \quad \hat{\Sigma} = \frac{1}{N} \{ \nu S + rI [S\hat{\beta}'\hat{\beta}S/\hat{\beta}S\hat{\beta}'] \}.$$

This expression is equal to the estimate (4.13) multiplied by  $1 - k^{-1} + N^{-1}$ . Therefore, since (4.13) converges in probability to the true value  $\Sigma$ , the estimate (8.18) converges in probability to  $(1 - k^{-1})\Sigma$ , and consequently, it is inconsistent. This fact is explained by the existence in the model (1.5) of an indefinitely increasing number of incidental parameters  $b_j$ .

The same inconsistency is found in linear regression analysis, *i.e.*, when we drop the restriction (1.1). The maximum likelihood estimate of  $\Sigma$  is then  $(\nu/N)S$ . When  $r$  tends to infinity,  $\nu/N \rightarrow 1 - k^{-1}$  and, therefore, since  $S$  is a consistent estimate of  $\Sigma$ , it follows that the maximum likelihood estimate of  $\Sigma$  in linear regression converges also to  $(1 - k^{-1})\Sigma$ . Obviously, this happens also in the ordinary univariate analysis of block designs, as was pointed out by J. Neyman and E. L. Scott ([22], Example 2) in the case of a block design with the same treatment applied to all experimental units.

**Acknowledgment.** The author wishes to acknowledge the valuable help of J. J. Schäffer and the referee, who have materially contributed to the final form of this paper.

#### REFERENCES

- [1] FORMAN S. ACTON, *Analysis of Straight-line Data*, John Wiley and Sons, New York, 1959.
- [2] R. J. ADCOCK, "Note on the method of least squares," *Analyst*, Vol. 4 (1877), pp. 183-184.

- [3] R. J. ADCOCK, "A problem in least squares," *Analyst*, Vol. 5 (1878), pp. 53-54.
- [4] R. L. ANDERSON AND T. A. BANCROFT, *Statistical Theory in Research*, McGraw Hill Book Co., New York, 1952.
- [5] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.
- [6] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946.
- [7] BERYL M. DENT, "On observation of points connected by a linear relation," *Proc. Physical Soc. London*, Vol. 47 (1935), pp. 92-108.
- [8] CAROLO FRIDERICO GAUSS, "Theoria combinationis observationum erroribus minimis obnoxiae," *Carl Friedrich Gauss Werke*, Vol. 4, pp. 1-108, Gottingen, 1880.
- [9] CAROLO FRIDERICO GAUSS, "Theoria motus corporum coelestium (1809)," *Carl Friedrich Gauss Werke*, Vol. 7, pp. 240-257, Leipzig, 1906.
- [10] PAUL R. HALMOS, *Finite Dimensional Vector Spaces*, 2nd ed., D. Van Nostrand, Princeton, N. J., 1958.
- [11] MAX HALPERIN, "Fitting of straight lines and prediction when both variables are subject to error," *Biometrics*, Vol. 15 (1959), p. 491.
- [12] G. W. HOUSNER AND J. F. BRENNAN, "The estimation of linear trends," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 380-388.
- [13] J. KIEFER AND J. WOLFOWITZ, "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 887-906.
- [14] T. KOOPMANS, *Linear Regression Analysis of Economic Time Series*, De Erven F. Bohn, Haarlem, 1937.
- [15] CHAS. H. KUMMELL, "Reduction of observed equations which contain more than one observed quantity," *Analyst*, Vol. 6 (1879), pp. 97-105.
- [16] D. V. LINDLEY, "Regression lines and the linear functional relationship," *Suppl. J. Roy. Stat. Soc.*, Vol. 9 (1947), pp. 218-244.
- [17] J. LUKOMSKI, "On some properties of multidimensional distributions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 236-246.
- [18] ALBERT MADANSKY, "The fitting of straight lines when both variables are subject to error," *J. Amer. Stat. Assn.*, Vol. 54 (1959), pp. 173-206.
- [19] WILLIAM G. MADOW, "Limiting distributions of quadratic and bilinear forms," *Ann. Math. Stat.*, Vol. 19 (1940), pp. 125-146.
- [20] H. B. MANN AND A. WALD, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 217-226.
- [21] J. NEYMAN, "Remarks on a paper by E. C. Rhodes," *J. Roy. Stat. Soc.*, Vol. 100 (1937), pp. 50-57.
- [22] J. NEYMAN AND ELIZABETH L. SCOTT, "Consistent estimates based on partially consistent observations," *Econometrica*, Vol. 16 (1948), pp. 1-32.
- [23] JOHN W. TUKEY, "Components in regression," *Biometrics*, Vol. 7 (1951), pp. 33-70.
- [24] M. J. VAN UVEN, "Adjustment of  $N$  points (in  $n$ -dimensional space) to the best linear ( $n - 1$ ) dimensional space," *Koninklijke Akademie van Wetenschappen te Amsterdam, Proceedings of the Section of Sciences*, Vol. 33 (1930), pp. 143-157 and 307-326.
- [25] CESÁREO VILLEGAS, "Sobre la estimación de una relación funcional lineal en el caso de un modelo a valores fijos," *Revista de la Unión Matemática Argentina y de la Asociación Física Argentina*, Vol. 18 (1958), pp. 179-180.

# SEQUENTIAL $\chi^2$ - AND $T^2$ -TESTS<sup>1</sup>

BY J. EDWARD JACKSON<sup>2</sup> AND RALPH A. BRADLEY<sup>3</sup>

*Virginia Polytechnic Institute*

**1. Summary.** Consider a multivariate normal population with mean  $\mu = (\mu_1, \dots, \mu_p)$  and covariance matrix  $\Sigma$ . Let  $\mu_0$  be a vector of constants,  $\bar{x}$  a vector of sample means based on  $n$  observations, and  $S_n$  the corresponding sample covariance matrix. The statistics considered are

$$(1.1) \quad \chi_n^2 = n(\bar{x} - \mu_0)\Sigma^{-1}(\bar{x} - \mu_0)'$$

and

$$(1.2) \quad T_n^2 = n(\bar{x} - \mu_0)S_n^{-1}(\bar{x} - \mu_0)'.$$

It is shown that probability-ratio tests for a sequential test of the composite hypothesis,

$$(1.3) \quad H_0: (\mu - \mu_0)\Sigma^{-1}(\mu - \mu_0)' = \lambda_0^2$$

against the alternative

$$(1.4) \quad H_1: (\mu - \mu_0)\Sigma^{-1}(\mu - \mu_0)' = \lambda_1^2$$

may be based on

$$(1.5) \quad p_{1n}/p_{0n} = [\exp - n(\lambda_1^2 - \lambda_0^2)/2] {}_0F_1(p/2; n\lambda_1^2\chi_n^2/4) / {}_0F_1(p/2; n\lambda_0^2\chi_n^2/4)$$

when  $\Sigma$  is known and

$$(1.6) \quad p_{1n}/p_{0n} = [\exp - n(\lambda_1^2 - \lambda_0^2)/2] {}_1F_1[n/2, p/2; n\lambda_1^2T_n^2/2(n-1+T_n^2)] / {}_1F_1[n/2, p/2; n\lambda_0^2T_n^2/2(n-1+T_n^2)]$$

when  $\Sigma$  is unknown and must be estimated from the sample. The sequential  $\chi^2$ -test is associated with (1.5) and the sequential  $T^2$ -test with (1.6).  ${}_0F_1$  and  ${}_1F_1$  are respectively forms of the generalized hypergeometric function  ${}_pF_q(a_1, \dots, a_p; c_1, \dots, c_q; x)$ , the second being the confluent hypergeometric function.

It is shown that the use of these probability ratios in sequential tests results in Type I and Type II errors of approximately  $\alpha$  and  $\beta$  when these values are used to obtain bounds on the probability ratios in the traditional way. It is also shown that the sequential tests terminate with probability unity. Bounds on the prob-

Received December 28, 1959; revised July 20, 1961.

<sup>1</sup> Research sponsored by the Statistics Branch, Office of Naval Research, U. S. Navy. Reproduction in whole or in part is permitted for any purpose of the United States Government.

<sup>2</sup> Now with the Eastman Kodak Company, Rochester, New York.

<sup>3</sup> Now with the Department of Statistics, Florida State University, Tallahassee, Florida.

ability ratios are translated into bounds on  $\chi_n^2$  and  $T_n^2$  themselves and tables have been prepared with more tables in preparation.

Procedures are also given to test sequentially whether or not two samples come from populations with the same means. The  $\chi^2$ -test is generalized to give simultaneous sequential tests on both the means and the covariance matrix.

The average sample number functions (ASN functions) are considered and approximations to them suggested. The operating characteristic functions (OC functions) are difficult to investigate and essentially are only known approximately at  $\lambda_0^2$  and  $\lambda_1^2$ .

**2. Introduction.** Modern techniques of sequential analysis were largely inspired by the work of Wald, summarized in his book [27]. This work was motivated by the need to cut down on the amount of work necessary in the acceptance sampling of military supplies.

Wald's procedures are based on a probability-ratio test and were largely developed for the test of a simple hypothesis against a simple alternative. For composite hypotheses, Wald proposed a *method of weight functions* but the method is cumbersome and no method of insuring optimum weights is available. Goldberg, as reported by Wallis [28], and Nandi [19] proposed a *method of frequency functions*. This method is now generally used in considering composite hypotheses and will be used here.

The method of frequency functions was used to develop the sequential  $t$ -test in the univariate case. This work was done independently by Rushton [24] [25] and Arnold [21]. We extend these methods to the multivariate problem to obtain the sequential  $T^2$ -test and also consider a sequential  $\chi^2$ -test. Applications are of consequence for, in the inspection of complex items, a number of characteristics are measured. Observations on these characteristics are often correlated and univariate sequential methods applied to each characteristic lead to confusion.

For the method of frequency functions, *observations* are successive values of the test-statistic and hence observations are no longer independent. Wald showed that the probability-ratio test could be used to obtain bounds on the test-statistic even though observations are not independent, but his work on termination, OC functions, and ASN functions no longer applies. Barnard [2] [3] and Cox [5] have independently established conditions under which the frequency function of a test statistic might be used in a sequential probability-ratio test and still guarantee approximately the risks  $\alpha$  and  $\beta$ . We quote Cox's theorem for later use.

**THEOREM.** "Let  $\mathbf{x} = [x_1, \dots, x_n]$  be random variables whose probability density function (p.d.f.) depends on unknown parameters  $\theta_1, \dots, \theta_p$ . The  $x_i$  themselves may be vectors. Suppose that

- (i)  $t_1, \dots, t_p$  are a functionally independent jointly sufficient set of estimators for  $\theta_1, \dots, \theta_p$ ;
- (ii) the distribution of  $t_1$  involves  $\theta_1$  but not  $\theta_2, \dots, \theta_p$ ;
- (iii)  $u_1, \dots, u_m$  are functions of  $\mathbf{x}$  functionally independent of each other and of  $t_1, \dots, t_p$ ;



(iv) there exists a set  $S$  of transformations of  $\mathbf{x} = [x_1, \dots, x_n]$  into  $\mathbf{x}^* = [x_1^*, \dots, x_n^*]$  such that

(a)  $t_1, u_1, \dots, u_m$  are unchanged by all transformations in  $S$ ;

(b) the transformation of  $t_2, \dots, t_p$  into  $t_2^*, \dots, t_p^*$  defined by each transformation in  $S$  is one-to-one;

(c) if  $T_2, \dots, T_p$  and  $T_2^*, \dots, T_p^*$  are two sets of values of  $t_2, \dots, t_p$  each having non-zero probability density under at least one of the distributions of  $\mathbf{x}$ , then there exists a transformation in  $S$  such that if  $t_2 = T_2, \dots, t_p = T_p$ , then  $t_2^* = T_2^*, \dots, t_p^* = T_p^*$ .

Then the joint p.d.f. of  $t_1, u_1, \dots, u_m$  factorizes into

$$g(t_1 | \theta_1) \ell(u_1, \dots, u_m, t_1),$$

where  $g$  is the p.d.f. of  $t_1$  and  $\ell$  does not involve  $\theta_1$ . The proof of this theorem is given in Cox's paper.

The application of the theorem is straight-forward. The theorem permits factorization of the sample p.d.f. in such a way that the probability ratio attempted from the sample p.d.f.'s under null and alternative hypotheses reduces to the probability ratio for the test statistic under null and alternative hypotheses. The composite hypotheses have been reduced to simple hypotheses on a single parameter involved in the distributions of the test statistics. In repetition,

$$(2.1) \quad p_{1n}/p_{0n} = g(t_{1n} | \theta_1) / g(t_{1n} | \theta_0)$$

in the notation of the theorem and where  $t_{1n}$  is the statistic  $t_1$  based on  $n$  observations. The sequential test is as follows:

- (i) Accept  $H_0$  if  $p_{1n}/p_{0n} \leq \beta/(1 - \alpha)$ .
- (ii) Accept  $H_1$  if  $p_{1n}/p_{0n} \geq (1 - \beta)/\alpha$ .
- (iii) Continue sampling if  $\beta/(1 - \alpha) < p_{1n}/p_{0n} < (1 - \beta)/\alpha$ .

Provided that the probability is one that the test terminates, the probabilities of error under the null and alternative hypotheses are approximately  $\alpha$  and  $\beta$  respectively (Wald [27], p. 43).

We test  $H_0$  of (1.3) against  $H_1$  of (1.4) using  $\chi_n^2$  in (1.1) or  $T_n^2$  in (1.2) depending on whether  $\Sigma$  is known or not. For the sequential  $\chi^2$ -test, the probability ratio is the ratio of two non-central  $\chi^2$ -densities as shown in (1.5). For the sequential  $T^2$ -test, the probability is the ratio of two non-central  $T^2$ -densities as shown in (1.6). In both (1.5) and (1.6) some simple combinations of terms have been made to obtain the forms shown.

**3. Fulfillment of the conditions of Cox's Theorem.** Verification of the conditions of Cox's Theorem has not been included by other authors using the theorem. Since we have not found the necessary verifications trivial either for this paper or others already published, we include a sketch of the required demonstrations.

Let  $\mathbf{x}$  be the vector  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  where  $\mathbf{x}_i$  is the  $i$ th observation vector on

$p$  multivariate normal variates. Then  $\mathbf{x}$  consists of  $n$  independent, equally distributed, multivariate normal observation vectors. The vector of variate means is  $\mathbf{y}$  and the dispersion matrix is  $\Sigma$  assumed known for the  $\chi^2$ -test and unknown for the  $T^2$ -test. The sample covariance matrix corresponding to  $\Sigma$  is  $\mathbf{S}_n$ . We may let  $\hat{\mathbf{x}}_i = \mathbf{x}_i - \mathbf{y}_0$  and  $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_n)$  and regard  $\hat{\mathbf{x}}$  as the original observation matrix of the Cox Theorem. Note that  $\mathbf{S}_n$  is invariant under such changes in location.

CONDITION (i). A vector of sufficient statistics for the elements of  $\mathbf{y}$  is  ${}_n\bar{\mathbf{x}}$  and for  $\hat{\mathbf{y}} = \mathbf{y} - \mathbf{y}_0$  is  ${}_n\hat{\mathbf{x}}$ ; the elements of  $\mathbf{S}_n$  and  ${}_n\hat{\mathbf{x}}$  are sufficient for  $\mathbf{y}$  and  $\Sigma$ . (Cf. Anderson [1], Sec. 3.3.3.) Define  $\mathbf{G}\mathbf{G}' = \Sigma^{-1}$  and  $\mathbf{E}\mathbf{E}' = \mathbf{S}_n^{-1}$ . We transform so that

$$(3.1) \quad {}_n\mathbf{y}' = n^{\frac{1}{2}}\Delta'({}_n\bar{\mathbf{x}} - \mathbf{y}_0)' = n^{\frac{1}{2}}\Delta'{}_n\hat{\mathbf{x}}'$$

and

$$(3.2) \quad \mathbf{n}' = \mathbf{G}'(\mathbf{y} - \mathbf{y}_0)' = \mathbf{G}'\hat{\mathbf{y}}'$$

with  $\Delta = \mathbf{G}$  for the  $\chi^2$ -test and  $\Delta = \mathbf{E}$  for the  $T^2$ -test. A spherical transformation is used to transform  ${}_n\mathbf{y}$  to  $\chi_n^2$  or  $T_n^2$ ,  $a_{1n}, \dots, a_{p-1,n}$ ,  $\chi_n^2$  or  $T_n^2 \geq 0, 0 \leq a_{in} \leq \pi$ ,  $i = 1, \dots, (p-2)$ ,  $0 \leq a_{p-1,n} \leq 2\pi$ . A similar transformation transforms  $\mathbf{n}$  to  $\lambda^2 = (\mathbf{y} - \mathbf{y}_0)\Sigma^{-1}(\mathbf{y} - \mathbf{y}_0)', \alpha_1, \dots, \alpha_{p-1}, \lambda^2 \geq 0, 0 \leq \alpha_i \leq \pi, i = 1, \dots, (p-2), 0 \leq \alpha_{p-1} \leq 2\pi$ . The transformations on  ${}_n\bar{\mathbf{x}}$  and  $\mathbf{y}$  and on  ${}_n\hat{\mathbf{x}}$  and  $\hat{\mathbf{y}}$  are one-to-one and it follows that  $\chi_n^2, a_{1n}, \dots, a_{p-1,n}$  are a set of sufficient statistics for  $\lambda^2, \alpha_1, \dots, \alpha_{p-1}$  and  $T_n^2, a_{1n}, \dots, a_{p-1,n}, \mathbf{S}_n$  for  $\lambda^2, \alpha_1, \dots, \alpha_{p-1}, \Sigma$ . We associate the sets of statistics with  $t_1, \dots, t_p$  in the theorem and the sets of parameters with  $\theta_1, \dots, \theta_p$  of the theorem; condition (i) follows.

CONDITION (ii). It was shown by Fisher [8] that the marginal distribution of  $\chi_n^2$  involves only  $\lambda^2$  and Hsu [14] and Bose and Roy [4] have the result for  $T_n^2$ . Condition (ii) is met when  $\lambda^2$  is associated with  $\theta_1$  and  $\chi_n^2$  or  $T_n^2$  with  $t_1$ .

CONDITION (iii). To consider the third condition of the theorem, we associate  $u_1, \dots, u_m$  with  $\chi_1^2, \dots, \chi_{n-1}^2$  or  $T_{p+1}^2, \dots, T_{n-1}^2$ . It is necessary to show that the statistics in the sets are functionally independent of each other and of  $\chi_n^2, a_{1n}, \dots, a_{p-1,n}$  or of  $T_n^2, a_{1n}, \dots, a_{p-1,n}, \mathbf{S}_n$ .

Condition (iii) seems intuitively obvious. Formal proof depends on a series of transformations. We rearrange the elements of  $\hat{\mathbf{x}}$  to yield a  $p$  by  $n$  matrix  $\hat{\mathbf{X}}'$  with  $(i, j)$  element  $\hat{x}_{ij} = x_{ij} - \mu_{i0}, i = 1, \dots, p, j = 1, \dots, n$ . Let

$$(3.3) \quad \mathbf{Y} = \mathbf{M}\hat{\mathbf{X}}'$$

where  $\mathbf{M}$  is the non-singular  $n$ -square matrix with  $j$ th row given by

$$[j^{-1}, \dots, j^{-1}, 0, \dots, 0],$$

the number of non-zero elements being  $j$ . The rows of  $\mathbf{Y}$  now depend on the sample means of the  $p$  variates, the  $j$ th row on the means of the first  $j$  observations,  $j = 1, \dots, n$ . At the next stage we transform  $\mathbf{Y}_j$ , the  $j$ th row of  $\mathbf{Y}$ , so that

$$(3.4) \quad \mathbf{Z}_j = \mathbf{Y}_j\Delta_j$$

where  $\Delta_j = \mathbf{G}$  for the  $\chi^2$ -test,  $j = 1, \dots, n$ , and  $\Delta_j = \mathbf{E}_j$ ,  $\mathbf{E}_j \mathbf{E}_j' = \mathbf{S}_j$ ,  $j = p+1, \dots, n$ , for the  $T^2$ -test;  $\mathbf{S}_j$  is the sample dispersion matrix similar to  $\mathbf{S}_n$  but based on the first  $j$  observation vectors. A spherical transformation is now applied to each  $\mathbf{Z}_j$ . For  $\mathbf{Z}_j$ , we obtain new variables  $\chi_j^2, a_{1j}, \dots, a_{p-1,j}$  with  $\chi_j^2 = \sum_{i=1}^p z_{ij}^2 = j, \bar{\mathbf{x}} \mathbf{\Sigma}^{-1} j \bar{\mathbf{x}}'$  or  $T_j^2, a_{1j}, \dots, a_{p-1,j}$  with  $T_j^2 = \sum_{i=1}^p z_{ij}^2 = j, \bar{\mathbf{x}} \mathbf{S}_j^{-1} j \bar{\mathbf{x}}'$ . For the  $\chi^2$ -test, through the series of non-singular, one-to-one transformations sketched here,  $\mathbf{x}$  or  $\bar{\mathbf{x}}$  has been transformed to a set of new variables such that  $\chi_1^2, \dots, \chi_n^2$  are functionally independent of each other and of  $\chi_n^2, a_{1n}, \dots, a_{p-1,n}$ . Similarly, for the  $T^2$ -test, each successive  $T_j^2$  is a function of one more row of  $\bar{\mathbf{X}}$  and  $T_n^2, a_{1n}, \dots, a_{p-1,n}$ ,  $\mathbf{S}_n$  all depend on all rows of  $\bar{\mathbf{X}}$ . Condition (iii) follows from these considerations.

CONDITION (iv). We rewrite Condition (iv) in terms of the present problems:

There exists a set of transformations  $\mathcal{S}$  of  $\mathbf{X}$  into  $\mathbf{X}^*$  such that:

- (a)  $\chi_1^2, \dots, \chi_n^2$  (or  $T_{p+1}^2, \dots, T_n^2$ ) are unchanged by all transformations in  $\mathcal{S}$ .
- (b) The transformation of  $a_{1n}, \dots, a_{p-1,n}$  (and  $\mathbf{S}_n$  for the  $T^2$ -test) into  $a_{1n}^*, \dots, a_{p-1,n}^*$  (and  $\mathbf{S}_n^*$ ) defined by each transformation in  $\mathcal{S}$  is one-to-one.
- (c) If  $A_1, \dots, A_{p-1}$ , (and  $\mathcal{C}_n$ ) and  $A_1^*, \dots, A_{p-1}^*$ , (and  $\mathcal{C}_n^*$ ) are two sets of values of  $a_{1n}, \dots, a_{p-1,n}$  (and  $\mathbf{S}_n$ ), each having non-zero probability density under at least one of the distributions of  $\mathbf{X}$ , there exists a transformation in  $\mathcal{S}$  such that, if  $a_{1n} = A_1, \dots, a_{p-1,n} = A_{p-1}$ , ( $\mathbf{S}_n = \mathcal{C}_n$ ), then  $a_{1n}^* = A_1^*, \dots, a_{p-1,n}^* = A_{p-1}^*$ , ( $\mathbf{S}_n = \mathcal{C}_n^*$ ).

The necessary classes of transformations are

$$(3.5) \quad \mathbf{X}^* = \mathbf{X} \mathbf{G} \mathbf{B} \mathbf{G}^{-1} \quad \text{or} \quad \mathbf{X}^* = \mathbf{X} \mathbf{E} \mathbf{B} \mathbf{C}'$$

respectively for the  $\chi^2$ - and  $T^2$ -tests where  $\mathbf{B}$  is any  $p$  by  $p$  orthogonal matrix and  $\mathbf{C}$  is any non-singular, triangular,  $p$  by  $p$  matrix.

We first consider the  $\chi^2$ -test. From (3.3) and (3.4),  $\mathbf{Z} = \mathbf{M} \bar{\mathbf{X}} \mathbf{G}$  and  $\mathbf{Z} \mathbf{B} = \mathbf{M} \bar{\mathbf{X}} \mathbf{G} \mathbf{B} = \mathbf{M} \bar{\mathbf{X}}^* \mathbf{G} = \mathbf{Z}^*$ . The transformation from  $\mathbf{Z}$  to  $\mathbf{Z}^*$  is orthogonal. Parts (a) and (b) of Condition (iv) follow at once. The sums of squares of elements in rows of  $\mathbf{Z}$  equal the corresponding sums of squares for  $\mathbf{Z}^*$  and are  $\chi_1^2, \dots, \chi_n^2$ . The transformation of  $\mathbf{Z}_n$  into  $\mathbf{Z}_n^*$  is one-to-one for each  $\mathbf{B}$  and, since  $\chi_n^2, a_{1n}, \dots, a_{p-1,n}$  follow from a spherical transformation on the elements of  $\mathbf{Z}_n$  and the corresponding transformation applies to  $\mathbf{Z}_n^*$ , we have a one-to-one transformation of  $a_{1n}, \dots, a_{p-1,n}$  to  $a_{1n}^*, \dots, a_{p-1,n}^*$ .

If  $A_1, \dots, A_{p-1}$  and  $A_1^*, \dots, A_{p-1}^*$  are two sets of values of  $a_{1n}, \dots, a_{p-1,n}$  suitably restricted between 0 and  $\pi$  or 0 and  $2\pi$  as the case may be, then  $\mathbf{Z}_n$  and  $\mathbf{Z}_n^*$  may be evaluated except for the scalar  $\chi_n$  which is the same in both cases. If these specified values yield  $\mathbf{Z}_n$  and  $\mathbf{Z}_n^*$ , they are related by  $\mathbf{Z}_n^* = \mathbf{Z}_n \mathbf{B}$  and this equation defines  $(p-1)$  independent equations on the elements of  $\mathbf{B}$ . There are also  $p(p+1)/2$  additional equations on the elements of  $\mathbf{B}$  imposed through the requirement that  $\mathbf{B}$  be orthogonal. The solution for the  $p^2$  elements of  $\mathbf{B}$  is not unique (except for  $p=2$ ) but matrices  $\mathbf{B}$  satisfying the requirements may be found and this is sufficient for (c) of Condition (iv).

For the  $T^2$ -test,  $\mathbf{Z}_j/j^{\frac{1}{2}} = j, \bar{\mathbf{x}}^* = j, \bar{\mathbf{x}} \mathbf{E} \mathbf{B} \mathbf{C}'$  and  $\mathbf{S}_j^* = \mathbf{C} \mathbf{B}' \mathbf{E}' \mathbf{S}_j \mathbf{E} \mathbf{B} \mathbf{C}'$ ,

$j = p + 1, \dots, n$ . Note that  $\mathbf{S}_n^* = \mathbf{C}\mathbf{C}'$  because  $\mathbf{E}'\mathbf{S}_n\mathbf{E} = \mathbf{I}$ . Part (a) of Condition (iv) follows since

$$\begin{aligned} T_j^{*2} &= j {}_j\tilde{\mathbf{x}}^* \mathbf{S}_j^{*-1} {}_j\tilde{\mathbf{x}}^* = j {}_j\tilde{\mathbf{x}} \mathbf{E} \mathbf{B} \mathbf{C}' \mathbf{C}^{-1} \mathbf{B}' \mathbf{E}^{-1} \mathbf{S}_j^{-1} \mathbf{E}'^{-1} \mathbf{B} \mathbf{C}^{-1} \mathbf{C} \mathbf{B}' \mathbf{E}_j' {}_j\tilde{\mathbf{x}}' \\ &= j {}_j\tilde{\mathbf{x}} \mathbf{S}_j^{-1} {}_j\tilde{\mathbf{x}}' = T_j^2, \quad j = p + 1, \dots, n. \end{aligned}$$

From (3.3), (3.4) and (3.5) it follows that  $\mathbf{Z}_n^* = \mathbf{Z}_n \mathbf{B}$ . Given  $\mathbf{B}$  and  $\mathbf{C}$ , the transformations of  $\mathbf{S}_n$  into  $\mathbf{S}_n^*$  and  $\mathbf{Z}_n$  into  $\mathbf{Z}_n^*$  are one-to-one and the transformation of  $\mathbf{Z}_n$  into  $\mathbf{Z}_n^*$  actually transforms  $T_n^2, a_{1n}, \dots, a_{p-1,n}$  into  $T_n^{*2}, a_{1n}^*, \dots, a_{p-1,n}^*$  with  $T_n^2 = T_n^{*2}$ . Part (b) is thus verified. In regard to Part (c), the existence of an appropriate  $\mathbf{B}$  is demonstrated exactly as for the  $\chi^2$ -test and  $\mathbf{C}$  is defined by  $\mathfrak{S}_n^*$  and  $\mathbf{E}$  by  $\mathfrak{S}_n$ . Hence the required transformation in  $\mathfrak{S}$  exists and Part (c) also follows for the  $T^2$ -test.

All of the conditions of the theorem have been fulfilled for both the  $\chi^2$ - and the  $T^2$ -tests. Hence the joint p.d.f.'s of  $\chi_1^2, \dots, \chi_n^2$  or  $T_{p+1}^2, \dots, T_n^2$  factor into  $g(\chi_n^2 | n\lambda^2) \ell(\chi_1^2, \dots, \chi_n^2)$  or  $g(T_n^2 | n\lambda^2) \ell(T_{p+1}^2, \dots, T_n^2)$  and  $p_{1n}/p_{0n}$  can be written as  $g(\chi_n^2 | n\lambda_1^2)/g(\chi_n^2 | n\lambda_0^2)$  or  $g(T_n^2 | n\lambda_1^2)/g(T_n^2 | n\lambda_0^2)$ .

The first of these is the ratio of two non-central  $\chi^2$ -densities with  $p$  degrees of freedom and non-centrality parameters  $n\lambda_1^2$  and  $n\lambda_0^2$  reducing to (1.5); the second is the ratio of two non-central  $T^2$ -densities with degrees of freedom  $p$  and non-centrality parameters  $n\lambda_1^2$  and  $n\lambda_0^2$  reducing to (1.6).

In many situations  $\lambda_0^2 = 0$  and then (1.5) reduces to

$$(3.6) \quad p_{1n}/p_{0n} = e^{-1/n\lambda_1^2} {}_0F_1(p/2; n\lambda_1^2\chi_n^2/4)$$

while (1.6) becomes

$$(3.7) \quad p_{1n}/p_{0n} = e^{-1/n\lambda_1^2} {}_1F_1[n/2, p/2; n\lambda_1^2 T_n^2/2(n-1+T_n^2)].$$

Furthermore, if  $p = 1$ , (3.6) reduces to

$$(3.8) \quad p_{1n}/p_{0n} = e^{-1/n\lambda_1^2} \cosh [\lambda_1 \sum_{j=1}^n (x_j - \mu_0)/\sigma]$$

and (3.7) to

$$(3.9) \quad p_{1n}/p_{0n} = e^{-1/n\lambda_1^2} {}_1F_1[n/2, \frac{1}{2}; n\lambda_1^2 t^2/2(n-1+t^2)].$$

Equation (3.8) is equation 9.5, page 135, in Wald's book; equation (3.9) is equation 5 given by Rushton [25] for the univariate sequential  $t$ -test.

A referee and W. J. Hall have pointed out that Cox's Theorem may be bypassed through use of an unpublished theorem of Stein and the fact that  $\chi_n^2$  and  $T_n^2$  are maximal invariants of sufficient statistics under certain groups of linear transformations implying that they are sufficient for any invariant statistics. This means that  $\chi_n^2$  is sufficient for  $(\chi_1^2, \dots, \chi_n^2)$  and  $T_n^2$  is sufficient for  $(T_{p+1}^2, \dots, T_n^2)$  and the factorizations resulting from Cox's Theorem are immediate. The reader is referred to Hall [12]. We have not relied on this approach

for reasons set forth at the beginning of this section and because the transformations involved are of interest in themselves.

**4. Termination proofs.** Let  $\chi_n^2$  and  $\bar{\chi}_n^2$  be the boundary values for  $\chi_n^2$  corresponding to  $p_{1n}/p_{0n} = \beta/(1 - \alpha)$  and  $p_{1n}/p_{0n} = (1 - \beta)/\alpha$  where  $p_{1n}/p_{0n}$  is defined in (1.5).

We assert that

$$P(\text{Sample Size} > n) \leq P(\chi_n^2 < \chi_n^2 < \bar{\chi}_n^2) = P_n$$

and proof of termination follows if we show that  $\lim_{n \rightarrow \infty} P_n = 0$ . This approach is similar to that given by Ray [23] who considered sequential analysis of variance.

Set  $U_n^2 = \chi_n^2/n$  and consider the corresponding limits  $U_n^2$  and  $\bar{U}_n^2$ . Erdelyi et al [7] shows that

$${}_0F_1(c; x) = e^{-2\sqrt{x}} {}_1F_1[(2c - 1)/2, 2c - 1; 4\sqrt{x}]$$

and, when this is applied in (1.5) and  $\chi_n^2$  replaced by  $nU_n^2$ , we have

$$p_{1n}/p_{0n} = f_n(U_n^2) = [\exp \{ \{-n(\lambda_1^2 - \lambda_0^2)/2\} - (n^2\lambda_1^2 U_n^2)^{\frac{1}{2}} + (n^2\lambda_0^2 U_n^2)^{\frac{1}{2}} \}] \frac{{}_1F_1[(p - 1)/2, p - 1; 2(n^2\lambda_1^2 U_n^2)^{\frac{1}{2}}]}{{}_1F_1[(p - 1)/2, p - 1; 2(n^2\lambda_0^2 U_n^2)^{\frac{1}{2}}]}$$

Intersections of the family of curves  $y = g_n(U^2) = \ln f_n(U^2)$  and of  $y = \ln [\beta/(1 - \alpha)]$  and  $y = \ln [(1 - \beta)/\alpha]$  determine  $U_n^2$  and  $\bar{U}_n^2$  respectively.

It can be shown that

$$g'_n(U^2) > 0 \quad \text{for } U^2 > 0$$

and

$$\lim_{n \rightarrow \infty} g'_n(U^2) = \infty.$$

It may be demonstrated also that

$$g_n(U^2) = n[ \{- (\lambda_1^2 - \lambda_0^2)/2\} + (U^2)^{\frac{1}{2}} [(\lambda_1^2)^{\frac{1}{2}} - (\lambda_0^2)^{\frac{1}{2}}] + O(1/n) ].$$

These results are straightforward, being based on the proposition that

$${}_1F_1(a, c; x) = [\Gamma(c)/\Gamma(a)] e^x x^{a-c} [1 + O(|x|^{-1})]$$

given by Erdelyi. It follows that  $y = g_n(U^2)$  defines a family of curves starting at  $y = -n(\lambda_1^2 - \lambda_0^2)/2$  for  $U^2 = 0$  and increasing strictly to  $+\infty$  as  $U^2 \rightarrow \infty$ . Hence  $g_n(U^2) = 0$  has one root and it is

$$U_0^2 = [\lambda_1^2 + 2(\lambda_1^2\lambda_0^2)^{\frac{1}{2}} + \lambda_0^2]/4 + O(1/n).$$

The intersection of the horizontal line  $y = \ln [\beta/(1 - \alpha)]$  with  $y = g_n(U^2)$  occurs where

$$- [(\lambda_1^2 - \lambda_0^2)/2] + (U^2)^{\frac{1}{2}} [(\lambda_1^2)^{\frac{1}{2}} - (\lambda_0^2)^{\frac{1}{2}}] + O(1/n) = n^{-1} \ln [\beta/(1 - \alpha)]$$

and hence

$$U_n^2 = [\lambda_1^2 + 2(\lambda_1^2 \lambda_0^2)^{1/2} + \lambda_0^2]/4 + O(1/n) = U_0^2 + O(1/n).$$

Similarly,

$$\bar{U}_n^2 = U_0^2 + O(1/n).$$

Consider  $U_n^2$ . This is a random variable that converges stochastically to  $\lambda^2 = (\mathbf{y} - \mathbf{y}_0) \Sigma^{-1} (\mathbf{y} - \mathbf{y}_0)'$  as  $n \rightarrow \infty$ . If  $\lambda^2 \neq U_0^2$ , the sequential process terminates with probability 1. If  $\lambda^2 = U_0^2$ , more powerful methods are required.

The work of David and Kruskal [6] suggested the following argument. Let  $P_T$  be the probability of termination of the sequential test.

$$P_T \geq 1 - P(\chi_n^2 \leq \chi^2 \leq \bar{\chi}_n^2) \geq 1 - (\bar{\chi}_n^2 - \chi_n^2) \sup_{(\chi_n^2, \bar{\chi}_n^2)} h(\chi^2; p, n\lambda^2)$$

where  $h(\chi^2; p, n\lambda^2)$  is the noncentral chi-square density with  $p$  degrees of freedom and parameter of noncentrality  $n\lambda^2$ . We show below that  $\lim_{n \rightarrow \infty} (\bar{\chi}_n^2 - \chi_n^2)$  is finite and  $\lim_{n \rightarrow \infty} \sup_{(\chi_n^2, \bar{\chi}_n^2)} h(\chi^2; p, n\lambda^2) = 0$  as may be shown by examination of that density. Hence  $P_T \geq \lim_{n \rightarrow \infty} [1 - P(\chi_n^2 \leq \chi^2 \leq \bar{\chi}_n^2)] = 1$  and  $P_T = 1$ . It remains to show that  $\lim_{n \rightarrow \infty} (\bar{\chi}_n^2 - \chi_n^2)$  is finite.

We use  $U_n^2$ ,  $\bar{U}_n^2$  and  $\bar{U}_n^2$  and recall that  $\bar{U}_n^2$  and  $U_n^2$  are such that  $g_n(U_n^2) = \ln [\beta/(1 - \alpha)]$  and  $g_n(\bar{U}_n^2) = \ln [\alpha/(1 - \beta)]$ . Consider  $g_n(U_n^2)$  in the interval  $(U_n^2, \bar{U}_n^2)$ . We apply the law of the mean in this interval and divide both sides by  $n$  to obtain

$$(4.1) \quad \frac{\ln \left( \frac{1 - \beta}{\beta} \cdot \frac{1 - \alpha}{\alpha} \right)}{n(\bar{U}_n^2 - U_n^2)} = \frac{g'_n[U_n^2 + \theta(\bar{U}_n^2 - U_n^2)]}{n}$$

where  $0 \leq \theta \leq 1$ .

$$g'_n(U^2) = \frac{1}{2} \left( \frac{n^2 \lambda_1^2}{U^2} \right)^{1/2} \left\{ \frac{{}_1F_1[(p+1)/2, p; 2(n^2 \lambda_1^2 U^2)^{1/2}]}{{}_1F_1[(p-1)/2, p-1; 2(n^2 \lambda_1^2 U^2)^{1/2}]} - 1 \right\} \\ - \frac{1}{2} \left( \frac{n^2 \lambda_0^2}{U^2} \right)^{1/2} \left\{ \frac{{}_1F_1[(p+1)/2, p; 2(n^2 \lambda_0^2 U^2)^{1/2}]}{{}_1F_1[(p-1)/2, p-1; 2(n^2 \lambda_0^2 U^2)^{1/2}]} - 1 \right\}.$$

But

$${}_1F_1(a, c; x) = [\Gamma(c)/\Gamma(a)] e^x x^{a-c} [1 + O(|x|^{-1})] \quad \text{and}$$

$$\frac{g'_n(U^2)}{n} = \frac{1}{2} \left[ \left( \frac{\lambda_1^2}{U^2} \right)^{1/2} - \left( \frac{\lambda_0^2}{U^2} \right)^{1/2} \right] + O\left(\frac{1}{n}\right).$$

Now

$$U_n^2, \bar{U}_n^2 \rightarrow U_0^2 = \left( \frac{(\lambda_1^2)^{1/2} + (\lambda_0^2)^{1/2}}{2} \right)^2$$

and in the neighborhood of  $U_0^2$  [in the interval  $(U_n^2, \bar{U}_n^2)$ ],



$$\frac{g'_n(U^2)}{n} = \frac{(\lambda_1^2)^{\frac{1}{2}} - (\lambda_0^2)^{\frac{1}{2}}}{(\lambda_1^2)^{\frac{1}{2}} + (\lambda_0^2)^{\frac{1}{2}}} + O(1/n)$$

and  $\lim_{n \rightarrow \infty} [g'_n(U^2)/n] > 0$  in the neighborhood when  $\lambda_1^2 > \lambda_0^2$  as required. Returning to (4.1), we find that the right-hand side has a finite limit and consequently  $n(\bar{U}_n^2 - U_n^2) = \bar{\chi}_n^2 - \chi_n^2$  has a finite limit.

The termination proof for the sequential  $\chi^2$ -test is now complete.

It is well known that the non-central  $T^2$ -distribution approaches the non-central  $\chi^2$ -distribution asymptotically with  $n$ . Then the argument above applies.  $\bar{U}_n^2$  and  $\bar{T}_n^2$ , the boundary values for the Sequential  $T^2$ -test, approach  $n[\lambda_1^2 + 2(\lambda_1^2\lambda_0^2)^{\frac{1}{2}} + \lambda_0^2]/4$  also.

**5. Two-sample cases.** The sequential techniques discussed can also be used for two-sample tests with paired observation vectors. Let the first population have mean vector  $\mathbf{u}^{(1)}$  and dispersion matrix  $\Sigma_{11}$ , the second  $\mathbf{u}^{(2)}$  and  $\Sigma_{22}$ . Suppose further that the cross-covariance matrix is  $\Sigma_{12}$ . Let  $\mathbf{u} = \mathbf{u}^{(1)} - \mathbf{u}^{(2)}$  and  $\mathbf{x}_i = \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$ ,  $i = 1, \dots, n$  where  $\mathbf{x}_i^{(1)}$  and  $\mathbf{x}_i^{(2)}$  are respectively the  $i$ th observation vectors for populations 1 and 2. The dispersion matrix of  $\mathbf{x}$  is  $\Sigma_{11} + \Sigma_{22} - \Sigma_{12} - \Sigma_{12}'$ . Now when  $\Sigma_{11}$ ,  $\Sigma_{22}$  and  $\Sigma_{12}$  are known, the two sample problem is reduced to an application of the Sequential  $\chi^2$ -test.

When the variance-covariance matrices are not known and must be estimated, the situation is even simpler. Again we define  $\mathbf{u} = \mathbf{u}^{(1)} - \mathbf{u}^{(2)}$  and use variates  $\mathbf{x} = \mathbf{x}^{(1)} - \mathbf{x}^{(2)}$ ,  $\Sigma_{11} + \Sigma_{22} - \Sigma_{12} - \Sigma_{12}'$  is estimated directly from the observation vectors  $\mathbf{x}_i = \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$ ,  $i = 1, \dots, n$  and this problem is reduced to that handled by the Sequential  $T^2$ -test.

**6. The ASN functions.** In the planning of sequential experiments information on the expected sample sizes ( $S$ ) is desirable. Wald [27] established approximate procedures for determining the ASN function when sequential observations are independent. These procedures have not been demonstrated to be valid when successive values, not independent, of a test statistic are considered.

Johnson [18] circumvented the problem by sampling additional groups rather than additional items within a group and hence considered successive independent values of a test statistic. This could be done for the Sequential  $\chi^2$ - and  $T^2$ -tests but usually would require too many observations.

Rushton [24, 25], in discussing the univariate sequential  $t$ -test, reduced the problem to a one-parameter problem by assuming that, after a number of observations, the variance was known. This led to lower bounds on the ASN function. This procedure does not seem applicable in the multivariate case because, even if  $\Sigma$  is assumed known, we are still dealing with composite hypotheses.

A third approach is the Monte Carlo technique. This has been used for the univariate sequential  $t$ -test and Freund and Appleby [9] have studied our tests.

A fourth method may be attempted. The method is based on the fact that, if one ignores excesses over the boundaries at the termination of a sequential test,

$$(6.1) \quad E[\ln(p_{1n}/p_{0n})] = (1 - \alpha)\ln[\beta/(1 - \alpha)] + \alpha\ln[(1 - \beta)/\alpha]$$

where  $H_0$  is true and

$$(6.2) \quad E[\ln(p_{1n}/p_{0n})] = \beta \ln[\beta/(1-\alpha)] + (1-\beta) \ln[(1-\beta)/\alpha]$$

where  $H_1$  is true. Now, in general,  $\ln(p_{1n}/p_{0n})$  will depend on  $n$  and a test statistic  $T_n$  based on the first  $n$  observations. What is needed is to express  $E[\ln(p_{1n}/p_{0n})]$  as a function of  $E(n)$ , the ASN number, (and of the parameters involved) and to then solve (6.1) and (6.2) for  $\varepsilon_0(n)$  and  $\varepsilon_1(n)$ , the required ASN numbers under  $H_0$  and  $H_1$  respectively. But a way of doing this has not been found for any sequential tests of composite hypotheses so far as can be ascertained by the authors. Bhate, in unpublished work, proposed approximating to  $E[\ln(p_{1n}/p_{0n})]$  by replacing  $T_n$  and  $n$  in  $\ln(p_{1n}/p_{0n})$  by  $E[T_n | n = E(n)]$ , the fixed sample-size expectation of  $T_n$  given that  $n = E(n)$ , and  $E(n)$  respectively. The expectation  $E[T_n | n = E(n)]$  is obtained under  $H_0$  for (6.1) and under  $H_1$  for (6.2). This procedure is seen intuitively to give a "central value" for the distribution of  $\ln(p_{1n}/p_{0n})$  and, upon appropriate substitutions in (6.1) and (6.2), to give equations in  $\varepsilon_0(n)$  and  $\varepsilon_1(n)$ , values of  $E(n)$  under  $H_0$  and  $H_1$  respectively, for solution. The method, crude though it may appear, has been used in a number of situations, for example, by Ray [22] for sequential analysis of variance and most recently by Hajnal [11] for a two-sample sequential  $t$ -test.

We have tried the method for the sequential  $\chi^2$ - and  $T^2$ -tests of this paper. Since  $E[\chi_n^2 | n = E(n)] = (p + n\lambda^2) |_{n=E(n)}$  and, using  $p_{1n}/p_{0n}$  in (1.5), we obtained the equations

$$(6.3) \quad \begin{aligned} & -\frac{1}{2}n(\lambda_1^2 - \lambda_0^2) + \ln {}_0F_1[p/2; n\lambda_1^2(p + n\lambda_0^2)/4] \\ & - \ln {}_0F_1[p/2; n\lambda_0^2(p + n\lambda_1^2)/4] \\ & = (1-\alpha) \ln[\beta/(1-\alpha)] + \alpha \ln[(1-\beta)/\alpha] \end{aligned}$$

and

$$(6.4) \quad \begin{aligned} & -\frac{1}{2}n(\lambda_1^2 - \lambda_0^2) + \ln {}_0F_1[p/2; n\lambda_1^2(p + n\lambda_1^2)/4] \\ & - \ln {}_0F_1[p/2; n\lambda_0^2(p + n\lambda_1^2)/4] \\ & = \beta \ln[\beta/(1-\alpha)] + (1-\beta) \ln[(1-\beta)/\alpha]. \end{aligned}$$

For brevity, these equations have been written in terms of  $n$  but solution of (6.3) for  $n$  yields  $\varepsilon_0(n)$  and of (6.4),  $\varepsilon_1(n)$ , both for the  $\chi^2$ -test. For the  $T^2$ -test, let  $x = T_n^2/2(n-1 + T_n^2)$  and note that

$$E[x | n = E(n)] = (n\lambda^2/2)\{1 - [\exp(-n\lambda^2/2)][(n-p)/n] \cdot {}_1F_1[n/2, (n+2)/2; n\lambda^2/2]\} |_{n=E(n)}$$

following Wishart [29]. With  $p_{1n}/p_{0n}$  in (1.6), the equations for the  $T^2$ -test corresponding with (6.3) and (6.4) are

$$(6.5) \quad \begin{aligned} & -\frac{1}{2}n(\lambda_1^2 - \lambda_0^2) + \ln {}_1F_1[n/2, p/2; n\lambda_1^2\varepsilon_0(x)] - \ln {}_1F_1[n/2, p/2; n\lambda_0^2\varepsilon_0(x)] \\ & = (1-\alpha) \ln[\beta/(1-\alpha)] + \alpha \ln[(1-\beta)/\alpha] \end{aligned}$$

and

$$(6.6) \quad -\frac{1}{2}n(\lambda_1^2 - \lambda_0^2) + \ln {}_1F_1[n/2, p/2; n\lambda_1^2\epsilon_1(x)] - \ln {}_1F_1[n/2, p/2; n\lambda_0^2\epsilon_1(x)] \\ = \beta \ln [\beta/(1 - \alpha)] + (1 - \beta) \ln [(1 - \beta)/\alpha]$$

where  $\epsilon_0(x) = \mathcal{E}[x | n = \mathcal{E}(n)]|_{\lambda^2=\lambda_0^2}$  and  $\epsilon_1(x) = \mathcal{E}[x | n = \mathcal{E}(n)]|_{\lambda^2=\lambda_1^2}$ .

Again we have left (6.5) and (6.6) in terms of  $n$  for simplicity but note that the solution of (6.5) for  $n$  yields  $\epsilon_0(n)$  and of (6.6),  $\epsilon_1(n)$ .

Solutions of (6.3) and (6.4) or of (6.5) and (6.6) do seem to provide the desired guides on the numbers of experimental units required for the planning of sequential experimentation. Solution of these equations can be accomplished iteratively with a high-speed computer and, since applications are likely to be repetitive, this need only be done initially in setting up an experimental or control program.

The principal justification for the Bhate method of approximating ASN numbers is that results agree sufficiently well with Monte Carlo studies for practical purposes. For verification of this, the reader is referred to the paper by Freund and Appleby. One other study, conducted by K. J. Arnold (Nat. Bur. Stds. [21]), is available for the sequential  $t$ -test with  $p = 1$ ,  $\lambda_0^2 = 0$ ,  $\lambda_1^2 = 1.0$  and  $\alpha = \beta = .05$ . In that study 500 sets of observations were sampled for the two values of  $\lambda^2$ ; the average sample size to reach a decision under  $H_0$  was 10.0 while the conjectural value was 10.7 and under  $H_1$  was 11.2 compared to 9.7. It is interesting to note that the actual  $\alpha$ - and  $\beta$ -values from this study were .044 and .034 respectively, somewhat different from the intended values. Ray [22] used this second example also but a rounding error occurs which makes his conjectured values appear to be closer to Arnold's results than they really are.

**7. Generalized  $\chi^2$ - and  $T^2$ -statistics.** In addition to the  $\chi^2$ - and  $T^2$ -statistics already discussed, there are two others in each case which deserve mention and complete the families of  $\chi^2$ - and  $T^2$ -statistics (Hotelling [13]). We adopt Hotelling's notation and also drop the subscript  $n$  used on  $\chi^2$ ,  $T^2$ ,  $\mathbf{S}$  and  $\bar{\mathbf{x}}$ . The  $\chi^2$ -test so far considered is  $\chi_M^2$  in Hotelling's notation and  $\Sigma$  becomes  $\Sigma_0$ . Now  $\lambda^2 = (\mathbf{y} - \mathbf{y}_0)\Sigma_0^{-1}(\mathbf{y} - \mathbf{y}_0)'$  becomes  $\lambda_M^2$  and the sequential  $\chi^2$ -test is for the hypotheses,

$$(7.1) \quad \begin{aligned} H_0 : \lambda_M^2 &= \lambda_{M_0}^2, \\ H_1 : \lambda_M^2 &= \lambda_{M_1}^2. \end{aligned}$$

The second statistic to be considered is

$$(7.2) \quad \chi_D^2 = (n - 1) \operatorname{tr} \mathbf{S} \Sigma_0^{-1}$$

for hypotheses

$$(7.3) \quad \begin{aligned} H_0 : \Sigma &= \Sigma_0 \quad (\text{or } \lambda_D^2 = \operatorname{tr} \Sigma \Sigma_0^{-1} = \lambda_{D_0}^2), \\ H_1 : \Sigma &> \Sigma_0 \quad (\text{or } \lambda_D^2 = \lambda_{D_1}^2 > \lambda_{D_0}^2). \end{aligned}$$

Usually we shall want  $\lambda_{D_0}^2 = p$  but this is not essential.  $\chi_D^2$  is distributed like  $\chi^2$  with  $(n-1)p$  degrees of freedom and noncentrality parameter  $(n-1)\lambda_D^2$ .  $\chi_M^2$  and  $\chi_o^2$  are multivariate extensions of univariate tests based on  $\bar{x}$  and  $s$ .

The sum  $\chi_o^2 = \chi_M^2 + \chi_D^2$  is a measure of the overall variation of the sample from standard.  $\chi_o^2$  is distributed like  $\chi^2$  with  $np$  degrees of freedom and parameter of noncentrality  $\lambda_o^2 = n\lambda_M^2 + (n-1)\lambda_D^2$ . An alternative form for  $\chi_o^2$  is

$$(7.4) \quad \chi_o^2 = \sum_{i=1}^n \chi_i^2$$

where  $\chi_i^2 = (\mathbf{x}_i - \mathbf{y}_0)'\Sigma_0^{-1}(\mathbf{x}_i - \mathbf{y}_0)$ ,  $i = 1, \dots, n$ .  $\chi_D^2$  could be obtained by subtraction,  $\chi_o^2 - \chi_M^2$ , and  $\mathbf{S}$  need not be computed. Logical hypotheses for use with  $\chi_o^2$  appear to be

$$(7.5) \quad \begin{aligned} H_0 : n\lambda_o^2 &= n\lambda_{M_0}^2 + (n-1)\lambda_{D_0}^2 = n\lambda_{o_0}^2 \\ H_1 : n\lambda_o^2 &= n\lambda_{M_1}^2 + (n-1)\lambda_{D_1}^2 = n\lambda_{o_1}^2. \end{aligned}$$

Sequential tests may be developed for (7.3) and (7.5) based on  $\chi_D^2$  and  $\chi_o^2$ . The probability-ratio statistics are respectively

$$p_{1n}/p_{0n} = e^{-(n-1)(\lambda_{D_1}^2 - \lambda_{D_0}^2)/2} \frac{{}_0F_1[(n-1)p/2; (n-1)\lambda_{D_1}^2 \chi_D^2/4]}{{}_0F_1[(n-1)p/2; (n-1)\lambda_{D_0}^2 \chi_D^2/4]}$$

and

$$p_{1n}/p_{0n} = e^{-n(\lambda_{o_1}^2 - \lambda_{o_0}^2)/2} \frac{{}_0F_1[np/2; n\lambda_{o_1}^2 \chi_o^2/4]}{{}_0F_1[np/2; n\lambda_{o_0}^2 \chi_o^2/4]}.$$

These sequential  $\chi^2$ -tests are developed just as the one based on  $\chi_M^2$  and would depend on the same set of tables for values of  $\underline{\chi}^2$  and  $\bar{\chi}^2$  except that often for  $\chi_M^2$  we would have  $\lambda_{M_0}^2 = 0$  and here tables are required for cases where neither null nor non-null values of  $\lambda^2$  are zero.

If the family of sequential  $\chi^2$ -test were used, say, in sampling inspection, the inspector could ascertain after each item inspected

- (i) whether or not the sample means differed significantly from standard,
- (ii) whether or not the variation about sample means was greater than the preassigned  $\Sigma_0$ , and
- (iii) whether or not the overall variability of the sample is larger than should have been expected.

Generalizations of the sequential  $T^2$ -test are not directly available; generalizations of the non-sequential  $T^2$ -test were developed.  $T^2$  in this paper corresponds to  $T_M^2$  in Hotelling's notation.  $T_D^2$  of Hotelling generally represents the variability in a subgroup of an experiment compared to, say, the average subgroup variability of an experiment. Rarely would such situations occur in sequential experimentation. Somewhat more conceivable is the situation where a sequential  $T_M^2$  test is run in parallel with a  $\chi_D^2$ -test, the test on variances is based on previous experience but the test on means depends only on the variability of the sample

itself.  $T_M^2$  and  $T_O^2$  are useful statistics in the multivariate analysis of variance and could perhaps be used in sequential multivariate schemes when more is known about the forms of their distributions. Sequential tests for the roots of determinantal equations might also prove useful and feasible but computational procedures would be difficult.

**8. Discussion.** We now discuss some problems that arise in using the sequential methods developed in this paper.

(i) *Tables.* Direct applications of our sequential procedures involve comparison of the probability ratio at each stage with  $\beta/(1-\alpha)$  and  $(1-\beta)/\alpha$ . This is laborious and requires evaluation of either  ${}_0F_1(c; x)$  or  ${}_1F_1(a, c; x)$  after each observation. Tables of both functions are available (Jackson [15], Nath [20], Rushton and Lang [26]) but Lagrangian interpolation of the logarithms of these functions is still necessary in most cases. It is better to prepare tables of the boundary values  $\chi_n^2$  and  $\bar{\chi}_n^2$  and  $T_n^2$  and  $\bar{T}_n^2$  so that only the test statistic is computed in applications. Tables now completed for  $\alpha = \beta = .05$  are given by Jackson and Bradley [16] and show  $\chi_n^2$ ,  $\bar{\chi}_n^2$ ,  $T_n^2$  and  $\bar{T}_n^2$  for  $p = 2$  (1) 9;  $\lambda_0^2 = 0$ ;  $\lambda_1^2 = .5, 1.0, 2.0$ ; maximum  $n$ : 60 for  $\lambda_1^2 = .5$ , 45 for  $\lambda_1^2 = 1.0$ , 30 for  $\lambda_1^2 = 2.0$ . R. J. Freund with Jackson at the Virginia Polytechnic Institute has completed some additional tables and a report [10] has been prepared. Publication of a separate volume of tables is contemplated when this work is complete.

(ii) *Determination of  $H_0$  and  $H_1$ .* Specifications of values of the noncentrality parameter  $\lambda^2$  lead to difficult administrative decisions. For sequential tests for means we would often take  $\lambda_0^2 = 0$  corresponding to  $\mu = \mu_0$ . Determination of  $\lambda_1^2$  is much more difficult in the multivariate case than for the univariate case since a  $p$ -dimensional ellipsoid related to problem specifications must be visualized. No single rule on specifying  $\lambda_1^2$  can be given and each problem has to be handled individually. Jackson and Bradley give some examples in connection with the sampling inspection of ballistic missiles and a paper showing these applications has been accepted for publication [17].

Sequential procedures should also be extended to cover the use of one-sided tolerances and essentially generalize the work of Goldberg (Wallis [28]).

(iii) *OC and ASN functions and truncation.* No explicit or even approximate expressions yet exist for the OC and ASN functions when the hypotheses under consideration are composite. Until such time as these expressions can be found, we must rely on Monte Carlo evaluations for a description of these properties. Little or no work has yet been done regarding truncation of sequential tests of composite hypotheses. Again, until such expressions are available, we must rely on Monte Carlo studies to show us the effect of truncation on the OC and ASN functions.

(iv) *Grouping.* These techniques were originally designed for the sampling of ballistic missiles, items which involve considerable expense. However, for a low-cost, high-volume process, sequential sampling by groups might be preferable to item-by-item sampling. Except for a few isolated cases like the binomial, no



optimum procedures have been worked out for sequential sampling by groups. The general procedure recommended by Wald, in our case, would be to take groups of say  $m$  observations per group and compare the resultant  $\chi_n^2$  or  $T_n^2$ , as the case may be, with the corresponding  $\chi_n^2$  and  $\bar{\chi}_n^2$  or  $T_n^2$  and  $\bar{T}_n^2$  where  $n$  is now equal to  $m$ ,  $2m$ ,  $3m$ ,  $\dots$ , etc. The effect of this procedure is to increase the average sample number and to decrease the size of  $\alpha$  and  $\beta$ . Except for empirical studies, the magnitudes of these changes are unknown but the directions of the changes are such that they compensate for each other to some extent.

## REFERENCES

- [1] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.
- [2] G. A. BARNARD, "The frequency justification of certain sequential tests," *Biometrika*, Vol. 39 (1952), pp. 144-150.
- [3] G. A. BARNARD, "The frequency justification of certain sequential tests—addendum," *Biometrika*, Vol. 40 (1953), pp. 468-469.
- [4] R. C. BOSE AND S. N. ROY, "The exact distribution of the studentized  $D^2$ -statistic," *Sankhya*, Vol. 4 (1938), pp. 19-38.
- [5] D. R. COX, "Sequential tests for composite hypotheses," *Proc. Camb. Philos. Soc.*, Vol. 48 (1952), pp. 290-299.
- [6] HERBERT T. DAVID AND WILLIAM H. KRUSKAL, "The WAGR sequential  $t$ -test reaches a decision with probability one," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 797-805.
- [7] ARTHUR ERDELYI, WILHELM MAGNUS, FRITZ OBENHETTINGER, AND FRANCESCO G. TRICOMI, *Higher Transcendental Functions*, Vol. 1, McGraw-Hill Book Co., New York, 1953.
- [8] R. A. FISHER, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc. London*, Vol. 121 (1928), pp. 654-673.
- [9] R. J. FREUND AND R. H. APPLEBY, "An empirical evaluation of multivariate sequential procedures for testing means," submitted to *Ann. Math. Stat.*
- [10] RUDOLF J. FREUND AND J. EDWARD JACKSON, "Tables to facilitate multivariate sequential tests for means," Technical Report No. 12 to Office of Naval Research, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, 1960.
- [11] J. HAJNAL, "A two-sample sequential  $t$ -test," *Biometrika*, Vol. 48 (1961), pp. 65-75.
- [12] WILLIAM JACKSON HALL, "On sufficiency and invariance with applications in sequential analysis," Institute of Statistics, University of North Carolina, Mimeo series No. 228, June, 1959.
- [13] HAROLD HOTELLING, "Multivariate quality control," *Techniques of Statistical Analysis*, edited by Churchill Eisenhart, Millard W. Hastay and W. Allen Wallis, McGraw-Hill Book Co., New York, 1947, pp. 111-184.
- [14] P. L. HSU, "Notes on Hotelling's generalized  $T$ ," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 231-243.
- [15] J. EDWARD JACKSON, "A short table of the hypergeometric function  ${}_2F_1(c; x)$ ," *Sankhya*, Vol. 22 (1960), pp. 351-356.
- [16] J. EDWARD JACKSON AND R. A. BRADLEY, "Multivariate sequential procedures for testing means," Technical Report No. 6 to Office of Naval Research, Virginia Polytechnic Institute, Blacksburg, 1959.
- [17] J. EDWARD JACKSON AND RALPH A. BRADLEY, "Sequential  $\chi^2$ - and  $T^2$ -tests and their application to an acceptance sampling problem," *Technometrics*, accepted for publication.
- [18] N. L. JOHNSON, "Sequential procedures in certain component of variance problems," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 357-366.



- [19] H. K. NANDI, "Use of well-known statistics in sequential analysis," *Sankhya*, Vol. 8 (1948), pp. 339-344.
- [20] P. NATH, "Confluent hypergeometric function," *Sankhya*, Vol. 11 (1951) pp. 155-166.
- [21] NATIONAL BUREAU OF STANDARDS, "Tables to facilitate sequential  $t$ -tests," *Applied Math. Series No. 7* (1951), prepared by K. J. Arnold.
- [22] W. D. RAY, "Sequential analysis applied to certain experimental designs in the analysis of variance," *Biometrika*, Vol. 43 (1956), pp. 388-403.
- [23] W. D. RAY, "A proof that the sequential probability ratio test (S.P.R.T.) of the general linear hypothesis terminates with probability unity," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 521-523.
- [24] S. RUSHTON, "On a sequential  $t$ -test," *Biometrika*, Vol. 37 (1950), pp. 326-333.
- [25] S. RUSHTON, "On a two-sided sequential  $t$ -test," *Biometrika*, Vol. 39 (1952), pp. 302-308.
- [26] S. RUSHTON AND E. D. LANG, "Tables of the confluent hypergeometric function," *Sankhya*, Vol. 13 (1954), pp. 377-411.
- [27] ABRAHAM WALD, *Sequential Analysis*, John Wiley and Sons, New York, 1947.
- [28] W. ALLEN WALLIS, "Use of variables in acceptance inspection for percent defective," *Techniques of Statistical Analysis*, edited by Churchill Eisenhart, Millard W. Hastay and W. Allen Wallis, McGraw-Hill Book Company, Inc., New York, 1947, pp. 3-93.
- [29] JOHN WISHART, "A note on the distribution of the correlation ratio," *Biometrika*, Vol. 24 (1932), pp. 441-456.

# ESTIMATING THE PARAMETERS OF NEGATIVE EXPONENTIAL POPULATIONS FROM ONE OR TWO ORDER STATISTICS

BY H. LEON HARTER

*Aeronautical Research Laboratory, Wright-Patterson Air Force Base*

**0. Summary.** This paper discusses the use of order statistics in estimating the parameters of (negative) exponential populations. For the one-parameter exponential population, the best linear unbiased estimators,  $\hat{\sigma}_k = c_k x_k$  and  $\hat{\sigma}_{lm} = c_l x_l + c_m x_m$ , of the parameter  $\sigma$  are given, based on one order statistic  $x_k$  and on two order statistics  $x_l$  and  $x_m$ . For samples of any size up through  $n = 100$ , a table is given of  $k$ ,  $l$ , and  $m$  and of the coefficients  $c_k$ ,  $c_l$ , and  $c_m$ , together with the coefficients of  $\sigma^2$  in the variances  $V_k$  and  $V_{lm}$  of the estimators, and the corresponding efficiencies  $E_k$  and  $E_{lm}$  (relative to the best linear unbiased estimator based on all order statistics). For the two-parameter exponential population, the best linear unbiased estimators,  $\hat{\alpha} = c_{al}x_l + c_{am}x_m$ ,  $\hat{\sigma} = c_{sl}x_l + c_{sm}x_m$ , and  $\hat{\mu} = c_{\mu l}x_l + c_{\mu m}x_m$ , of the parameters  $\alpha$  and  $\sigma$  and the mean  $\mu = \alpha + \sigma$  are given, based on two order statistics  $x_l$  and  $x_m$ . For samples of any size up through  $n = 100$ , a table is given of  $m$  ( $l$  is always 1 for the best estimator) and of factors  $c_\alpha$  and  $c_\sigma$  for computing the coefficients  $c_{al} = 1 + c_\alpha$ ,  $c_{am} = -c_\alpha$ ,  $c_{sl} = -c_\sigma$ ,  $c_{sm} = c_\sigma$ ,  $c_{\mu l} = 1 + c_\alpha - c_\sigma$ , and  $c_{\mu m} = c_\sigma - c_\alpha$ , together with the coefficients of  $\sigma^2$  in the variances  $V_{\hat{\alpha}}$ ,  $V_{\hat{\sigma}}$ , and  $V_{\hat{\mu}}$  of the estimators, and the corresponding efficiencies  $E_{\hat{\alpha}}$ ,  $E_{\hat{\sigma}}$ , and  $E_{\hat{\mu}}$  (relative to the best linear unbiased estimators based on all order statistics).

**1. Introduction.** Since the publication, in 1946 and 1948 respectively, of papers by Mosteller [4] and by Wilks [10], a great deal of attention has been given to the use of order statistics in various statistical procedures, including the estimation of the parameters of various populations. Among the first to use this method for exponential populations was Halperin [2] in 1952. Since that time, Epstein and Sobel [1], Sarhan [5], [6], Sarhan and Greenberg [7], [8], and Sarhan, Greenberg, and Ogawa [9] have discussed various aspects of the use of order statistics in the estimation of the parameters of exponential populations. Most of these authors have considered best linear unbiased estimators based on all order statistics, or, in the case of truncated or censored samples, on all available order statistics. The last of the above papers includes simplified estimators based on two order statistics, with tables for samples of any size up through  $n = 20$ . The present paper contains more accurate tables for samples of any size up through  $n = 100$ , not only for estimators based on two order statistics, but also, in the case of the one-parameter exponential population, for estimators based on one order statistic.

In a previous paper, the author [3] studied estimators of the standard deviation

Received April 3, 1961; revised June 12, 1961.

of normal, rectangular, and one-parameter exponential populations based on sample ranges and quasi-ranges. The results were quite satisfactory for the symmetric populations, but not for the one-parameter exponential, for which it was found that estimators based on a single order statistic are more efficient. That discovery led to the further research reported in this paper.

## 2. Estimators of $\sigma$ for the One-Parameter Exponential Population.

2.1. *Estimators based on one order statistic.* For the one-parameter exponential population with parameter  $\sigma$ , which is both the mean and the standard deviation of the population, the probability density function  $f_1(x)$  is equal to  $(1/\sigma)\exp(-x/\sigma)$  for  $0 \leq x < \infty$ , and zero elsewhere. The sample mean  $\bar{x}$ , which has variance  $\sigma^2/n$ , is the minimum variance unbiased estimator, and also the maximum likelihood estimator, of the parameter  $\sigma$ . The expected value and the variance of the  $k$ th order statistic,  $x_k$ , of a sample of size  $n$  from this population are given (see Epstein and Sobel [1]) by

$$(1) \quad E(x_k) = \sigma \sum_{i=1}^k a_i$$

and

$$(2) \quad \text{var } x_k = \sigma^2 \sum_{i=1}^k a_i^2,$$

where  $a_i = 1/(n - i + 1)$ . An unbiased estimator of the parameter  $\sigma$ , based on the order statistic  $x_k$ , is given by  $\bar{\sigma}_k = c_k x_k$ , where

$$(3) \quad c_k = 1 / \sum_{i=1}^k a_i.$$

The variance of this estimator is given by

$$(4) \quad V_k = \sigma^2 \sum_{i=1}^k a_i^2 / \left( \sum_{i=1}^k a_i \right)^2,$$

and its efficiency (relative to the minimum variance unbiased estimator  $\bar{x}$ ) is  $E_k = \text{var } \bar{x} / V_k$ , where, as mentioned above,  $\text{var } \bar{x} = \sigma^2/n$ . Thus the relative efficiency  $E_k$  is given by

$$(5) \quad E_k = \left( \sum_{i=1}^k a_i \right)^2 / \left( n \sum_{i=1}^k a_i^2 \right).$$

The best estimator of  $\sigma$ , based on one order statistic  $x_k$ , is the one for that value of  $k$  which minimizes  $V_k$  (maximizes  $E_k$ ). The author is not aware of any analytical method for determining the value of  $k$  which yields the best estimator of  $\sigma$ ; hence, for each value of  $n$ ,  $V_k$  was computed for  $k = 1(1)n$ . When the best value of  $k$  for a given  $n$  had been found, the corresponding  $c_k$  and  $E_k$  were also computed. The computations were performed on the IBM 1620 computer. Table 1 gives, for  $n = 1(1)100$ , the value of  $k$  for the best estimator of  $\sigma$ , the coefficient  $c_k$  (to 6 significant figures), the coefficient,  $V_k/\sigma^2$ , of  $\sigma^2$  in the variance  $V_k$  of the

estimator (to 7 significant figures or 6 decimal places, whichever is less accurate), and the relative efficiency  $E_k$  (to 5 significant figures). The tabular values of  $c_k$ ,  $V_k/\sigma^2$ , and  $E_k$  are accurate to within a unit in the last place given.

2.2. *Estimators based on two order statistics.* Unbiased linear estimators of the parameter  $\sigma$ , based on two order statistics  $x_l$  and  $x_m$ , are given by  $\bar{\sigma}_{lm} = c_l x_l + c_m x_m$ , where  $c_l E(x_l) + c_m E(x_m) = \sigma$ , with  $E(x_l)$  and  $E(x_m)$  given by equation (1), if  $k$  takes the values  $l$  and  $m$ . The variance of such an estimator is given by

$$(6) \quad V_{lm} = c_l^2 \text{var } x_l + c_m^2 \text{var } x_m + 2c_l c_m \text{cov}(x_l, x_m),$$

where  $\text{var } x_l$  and  $\text{var } x_m$  are given by equation (2), if  $k$  takes the values  $l$  and  $m$ , and  $\text{cov}(x_l, x_m)$  is given (see Sarhan [5]) by

$$(7) \quad \text{cov}(x_l, x_m) = \sigma^2 \sum_1^l a_i^2 = \text{var } x_l, \quad (l < m).$$

It can be shown that, for given values of  $l$  and  $m$ , the values of  $c_l$  and  $c_m$  which yield the unbiased estimator  $\bar{\sigma}_{lm}$  with minimum variance are

$$(8) \quad c_l = 1 / \left( \sum_1^l a_i + \lambda \sum_1^m a_i \right)$$

and

$$(9) \quad c_m = \lambda c_l,$$

where

$$(10) \quad \lambda = \sum_{l+1}^m a_i \sum_1^l a_i^2 / \left( \sum_1^l a_i \sum_1^m a_i^2 - \sum_1^m a_i \sum_1^l a_i^2 \right).$$

By substituting from equations (2) and (7)–(9) into equation (6), one finds that the minimum variance of  $\bar{\sigma}_{lm}$ , for given  $l$  and  $m$ , is

$$(11) \quad V_{lm} = \sigma^2 \left[ (1 + 2\lambda) \sum_1^l a_i^2 + \lambda^2 \sum_1^m a_i^2 \right] / \left( \sum_1^l a_i + \lambda \sum_1^m a_i \right)^2.$$

The efficiency of this estimator (relative to the minimum variance unbiased estimator  $\bar{x}$ ) is  $E_{lm} = \text{var } \bar{x} / V_{lm}$ , where, as before,  $\text{var } \bar{x} = \sigma^2/n$ . Thus the relative efficiency  $E_{lm}$  is given by

$$(12) \quad E_{lm} = \left( \sum_1^l a_i + \lambda \sum_1^m a_i \right)^2 / n \left[ (1 + 2\lambda) \sum_1^l a_i^2 + \lambda^2 \sum_1^m a_i^2 \right].$$

The best estimator of  $\sigma$ , based on two order statistics  $x_l$  and  $x_m$ , is the one for those values of  $l$  and  $m$  which minimize  $V_{lm}$  (maximize  $E_{lm}$ ). The author is not aware of any analytical method for determining these values of  $l$  and  $m$ ; hence, for each value of  $n$ ,  $V_{lm}$  was computed for  $l = 1(1)(n-1)$  and  $m = (l+1)(1)n$ . When the best values of  $l$  and  $m$  for a given  $n$  had been found, the corresponding  $c_l$ ,  $c_m$ , and  $E_{lm}$  were also computed. The computations were performed on the IBM 1620 computer. Table 1 gives, for  $n = 2(1)100$ , the values of  $l$  and

$m$  for the best estimator of  $\sigma$ , the coefficients  $c_l$  and  $c_m$  (to 6 significant figures), the coefficient,  $V_{lm}/\sigma^2$ , of  $\sigma^2$  in the variance  $V_{lm}$  of the estimator (to 7 decimal places), and the relative efficiency  $E_{lm}$  (to 5 significant figures). The tabular values of  $c_l$ ,  $c_m$ ,  $V_{lm}/\sigma^2$ , and  $E_{lm}$  are accurate to within a unit in the last place given.

**3. Estimators or Parameters  $\alpha$ ,  $\sigma$  and Mean  $\mu$  for the Two-Parameter Exponential Population from Two Order Statistics.** For the two-parameter exponential population with parameters  $\alpha$  and  $\sigma$ , the probability density function  $f_2(x)$  is  $(1/\sigma)\exp[-(x-\alpha)/\sigma]$  for  $\alpha \leq x < \infty$ , and zero elsewhere. The mean of this population, which will be denoted by  $\mu$ , is equal to  $\alpha + \sigma$ . For a sample of size  $n$  from this population, the expected value of the  $k$ th order statistic,  $x_k$ , exceeds by  $\alpha$  the value given in equation (1) for the one-parameter exponential population, and thus is given by

$$(13) \quad E(x_k) = \alpha + \sigma \sum_{i=1}^k a_i.$$

The variance of  $x_k$  and the covariance of  $x_l$  and  $x_m$  are the same as for the one-parameter exponential population, and hence are given by equations (2) and (7), respectively.

Unbiased linear estimators of the parameters  $\alpha$  and  $\sigma$  and the mean  $\mu$  may be obtained from any two order statistics  $x_l$  and  $x_m$ . These estimators are of the form

$$(14) \quad \bar{\alpha} = c_{\alpha l}x_l + c_{\alpha m}x_m,$$

$$(15) \quad \bar{\sigma} = c_{\sigma l}x_l + c_{\sigma m}x_m,$$

and

$$(16) \quad \bar{\mu} = c_{\mu l}x_l + c_{\mu m}x_m.$$

It has been shown (see Sarhan, Greenberg, and Ogawa [9]) that, for given  $l$  and  $m$ , the coefficients in the best linear estimators based on two order statistics  $x_l$  and  $x_m$  are given by

$$(17) \quad c_{\alpha l} = 1 + c_{\alpha}, c_{\alpha m} = -c_{\alpha},$$

$$(18) \quad c_{\sigma l} = -c_{\sigma}, c_{\sigma m} = c_{\sigma},$$

and

$$(19) \quad c_{\mu l} = 1 + c_{\alpha} - c_{\sigma}, c_{\mu m} = c_{\sigma} - c_{\alpha},$$

where

$$(20) \quad c_{\alpha} = \sum_{i=1}^l a_i / \sum_{i=1}^m a_i$$

and

$$(21) \quad c_{\sigma} = 1 / \sum_{i=1}^m a_i.$$

The variance of the estimator  $\bar{\sigma}$  is given by

$$(22) \quad V_{\bar{\sigma}} = \sigma^2 \sum_{i=1}^m a_i^2 / \left( \sum_{i=1}^m a_i \right)^2,$$

and the variances of the estimators  $\bar{\alpha}$  and  $\bar{\mu}$  are given by

$$(23) \quad \begin{aligned} V_{\bar{\alpha}} &= \sigma^2 \sum_{i=1}^l a_i^2 + \left( \sum_{i=1}^l a_i \right)^2 V_{\bar{\sigma}} \\ &= \sigma^2 \left[ \sum_{i=1}^l a_i^2 + \left( \sum_{i=1}^l a_i \right)^2 \sum_{i=1}^m a_i^2 / \left( \sum_{i=1}^m a_i \right)^2 \right] \end{aligned}$$

and

$$(24) \quad \begin{aligned} V_{\bar{\mu}} &= \sigma^2 \sum_{i=1}^l a_i^2 + \left( \sum_{i=1}^l a_i - 1 \right)^2 V_{\bar{\sigma}} \\ &= \sigma^2 \left[ \sum_{i=1}^l a_i^2 + \left( \sum_{i=1}^l a_i - 1 \right)^2 \sum_{i=1}^m a_i^2 / \left( \sum_{i=1}^m a_i \right)^2 \right]. \end{aligned}$$

The best linear unbiased estimators of  $\alpha$ ,  $\sigma$ , and  $\mu$  based on all order statistics (see Sarhan and Greenberg [7]) are

$$(25) \quad \hat{\alpha} = \left[ (n^2 - 1)x_1 - \sum_{i=2}^n x_i \right] / [n(n-1)],$$

$$(26) \quad \hat{\sigma} = \left[ \sum_{i=2}^n x_i - (n-1)x_1 \right] / (n-1),$$

and

$$(27) \quad \hat{\mu} = \sum_{i=1}^n x_i / n = \bar{x}.$$

These are also the maximum likelihood estimators. Their variances are

$$(28) \quad V_{\hat{\alpha}} = \sigma^2 / [n(n-1)],$$

$$(29) \quad V_{\hat{\sigma}} = \sigma^2 / (n-1),$$

and

$$(30) \quad V_{\hat{\mu}} = \sigma^2 / n.$$

The efficiencies of the estimators  $\bar{\alpha}$ ,  $\bar{\sigma}$ , and  $\bar{\mu}$  (relative to the best linear unbiased estimators  $\hat{\alpha}$ ,  $\hat{\sigma}$ , and  $\hat{\mu}$  based on all order statistics) are given by

$$(31) \quad E_{\bar{\alpha}} = V_{\hat{\alpha}} / V_{\bar{\alpha}} = \left( \sum_{i=1}^m a_i \right)^2 / n(n-1) \cdot \left[ \sum_{i=1}^l a_i^2 \left( \sum_{i=1}^m a_i \right)^2 + \left( \sum_{i=1}^l a_i \right)^2 \sum_{i=1}^m a_i^2 \right],$$

$$(32) \quad E_{\bar{\sigma}} = V_{\hat{\sigma}} / V_{\bar{\sigma}} = \left( \sum_{i=1}^m a_i \right)^2 / (n-1) \sum_{i=1}^m a_i^2,$$



and

$$(33) \quad E_{\hat{\mu}} = V_{\hat{\mu}}/V_{\hat{\sigma}} = \left( \sum_{i=1}^m a_i \right)^2 / n \left[ \sum_{i=1}^l a_i^2 \left( \sum_{i=1}^m a_i \right)^2 + \left( \sum_{i=1}^l a_i - 1 \right)^2 \sum_{i=1}^m a_i^2 \right].$$

The best estimators of  $\alpha$ ,  $\sigma$ , and  $\mu$ , based on two order statistics  $x_l$  and  $x_m$ , are those for the values of  $l$  and  $m$  which minimize  $V_{\hat{\alpha}}$ ,  $V_{\hat{\sigma}}$ , and  $V_{\hat{\mu}}$  (maximize  $E_{\hat{\alpha}}$ ,  $E_{\hat{\sigma}}$ , and  $E_{\hat{\mu}}$ ). It can be shown that, for a fixed value of  $m$ , the variances of the estimators are smallest when  $l = 1$ . It can be seen from equations (23) and (24) that, for a fixed value of  $l$ , the value of  $m$  which minimizes  $V_{\hat{\sigma}}$  also minimizes  $V_{\hat{\alpha}}$  and  $V_{\hat{\mu}}$ . The author is not aware of any purely analytical method of determining the best value of  $m$ ; hence, for each value of  $n$ ,  $V_{\hat{\sigma}}$  was computed for  $l = 1$  and  $m = 2(1)n$ . When the best value of  $m$  for a given  $n$  had been found, the corresponding  $c_{\alpha}$ ,  $c_{\sigma}$ ,  $V_{\hat{\alpha}}$ ,  $V_{\hat{\mu}}$ ,  $E_{\hat{\alpha}}$ ,  $E_{\hat{\sigma}}$ , and  $E_{\hat{\mu}}$  were also computed. The computations were performed on the IBM 1620 computer. Table 2 gives, for  $n = 2(1)100$ , the value of  $m$  for the best estimators of  $\alpha$ ,  $\sigma$ , and  $\mu$ , the factors  $c_{\alpha}$  and  $c_{\sigma}$  (to 6 significant figures or 6 decimal places, whichever is less accurate), the coefficient,  $V_{\hat{\alpha}}/\sigma^2$ , of  $\sigma^2$  in the variance  $V_{\hat{\alpha}}$  (to 7 significant figures or 9 decimal places, whichever is less accurate), the coefficients,  $V_{\hat{\sigma}}/\sigma^2$  and  $V_{\hat{\mu}}/\sigma^2$ , of  $\sigma^2$  in the variances  $V_{\hat{\sigma}}$  and  $V_{\hat{\mu}}$  (to 7 significant figures or 7 decimal places, whichever is less accurate), and the relative efficiencies  $E_{\hat{\alpha}}$ ,  $E_{\hat{\sigma}}$ , and  $E_{\hat{\mu}}$  (to 5 significant figures). The tabular values of  $c_{\alpha}$ ,  $c_{\sigma}$ ,  $V_{\hat{\alpha}}/\sigma^2$ ,  $V_{\hat{\sigma}}/\sigma^2$ ,  $V_{\hat{\mu}}/\sigma^2$ ,  $E_{\hat{\alpha}}$ ,  $E_{\hat{\sigma}}$ , and  $E_{\hat{\mu}}$  are accurate to within a unit in the last place given.

#### 4. Remarks.

(i) The variance of the best estimator of  $\sigma$  for the two-parameter exponential population based on two order statistics  $x_1$  and  $x_m$  from a sample of size  $n$  is the same as the variance of the best estimator of  $\sigma$  for the one-parameter exponential population based on one order statistic  $x_k$  from a sample of size  $n - 1$ , with  $k = m - 1$ . This can be seen by a comparison of equations (4) and (22), though the author did not observe this fact until confronted with equal numerical values. The relative efficiencies of these estimators are also equal, since in each case the variance of the best linear unbiased estimator based on all order statistics is  $\sigma^2/(n - 1)$ .

(ii) For the one-parameter exponential population, the values  $k = 80$ ,  $l = 64$ , and  $m = 93$ , with coefficients  $c_k = 0.629074$ ,  $c_l = 0.522657$ , and  $c_m = 0.181399$  and relative efficiencies  $E_k = 65.093\%$  and  $E_{lm} = 82.460\%$ , for  $n = 100$  may be compared with the results obtained by Sarhan, Greenberg, and Ogawa [9], whose corresponding asymptotic values are  $0.7968n$ ,  $0.6386n$ ,  $0.9266n$ ,  $0.6275$ ,  $0.5232$ ,  $0.1790$ ,  $64.76\%$ , and  $82.03\%$ .

(iii) For the two-parameter exponential population, it can be seen from equations (20) and (21) that, since  $l = 1$  for the best estimators,  $c_{\alpha} = a_1 c_{\sigma} = c_{\sigma}/n$ . For convenience, however, separate columns for  $c_{\alpha}$  and  $c_{\sigma}$  are given in Table 2.

## REFERENCES

- [1] EPSTEIN, BENJAMIN AND SOBEL, MILTON, "Life testing," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 486-502.
- [2] HALPERIN, MAX, "Maximum likelihood estimation in truncated samples," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 226-238.
- [3] HARTER, H. LEON, "The use of sample quasi-ranges in estimating population standard deviation," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 980-999.
- [4] MOSTELLER, FREDERICK, "On some useful 'inefficient' statistics," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 377-408.
- [5] SARHAN, A. E., "Estimation of the mean and standard deviation by order statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 317-328.
- [6] SARHAN, A. E., "Estimation of the mean and standard deviation by order statistics. Part III," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 576-592.
- [7] SARHAN, A. E. AND GREENBERG, B. G., "Tables for best linear estimates by order statistics of the parameters of single exponential distributions from singly and doubly censored samples," *J. Amer. Stat. Assn.*, Vol. 52 (1957), pp. 58-87.
- [8] SARHAN, A. E. AND GREENBERG, B. G., "Estimation problems in the exponential distribution using order statistics," *Proc. Stat. Techniques in Missile Evaluation Symposium*, Virginia Polytechnic Institute (1958), pp. 123-175.
- [9] SARHAN, A. E., GREENBERG, B. G., AND OGAWA, JUNJIRO, "Simplified estimates for the exponential distribution," Interim Technical Report No. 1, Grant No. DA-ORD-7, University of North Carolina, 1960.
- [10] WILKS, S. S., "Order statistics," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 6-50.

TABLE 1  
Best Estimators of Parameter  $\mu$  of 1-Parameter Exponential Population

$n$	From 1 Order Statistic, $x_k$				From 2 Order Statistics, $x_l$ and $x_m$					
	$k$	$c_k$	$V_k/\sigma^2$	$E_k(\%)$	$l$	$m$	$c_l$	$c_m$	$V_{lm}/\sigma^2$	$E_{lm}(\%)$
1	1	1.00000	1.000000	100.00						
2	2	.666667	.5555556	90.000	1	2	.500000	.500000	.5000000	100.00
3	3	.545455	.4049587	82.313	2	3	.447368	.342105	.3421053	97.436
4	4	.480000	.3280000	76.220	3	4	.413043	.265217	.2652174	94.262
5	5	.437956	.2807289	71.243	3	5	.527997	.256818	.2140154	93.451
6	5	.689655	.2337165	71.311	4	6	.493892	.216654	.1805452	92.313
7	6	.627803	.2017178	70.820	5	7	.467528	.188618	.1571815	90.887
8	7	.582121	.1787245	69.940	5	8	.553626	.187760	.1393976	89.672
9	8	.546756	.1613595	68.859	6	9	.527062	.167990	.1247200	89.088
10	8	.699806	.1468046	68.118	7	10	.505032	.152501	.1132201	88.324
11	9	.657948	.1333457	68.175	8	11	.486391	.140031	.1039622	87.444
12	10	.623748	.1225454	68.002	8	12	.552632	.140623	.0960924	86.722
13	11	.595191	.1136773	67.668	9	13	.533452	.130469	.0891540	86.281
14	12	.570919	.1062578	67.222	10	14	.516702	.121903	.0833001	85.748
15	12	.673448	.0994728	67.020	11	15	.501916	.114575	.0782930	85.150
16	13	.646247	.0932310	67.038	10	15	.519329	.217012	.0735428	84.984
17	14	.622580	.0878686	66.945	11	16	.506584	.204426	.0692778	84.910
18	15	.601766	.0832093	66.766	12	17	.495052	.193425	.0655497	84.753
19	16	.583292	.0791212	66.520	12	18	.531320	.193368	.0621307	84.711
20	16	.660325	.0752377	66.456	13	19	.519816	.183870	.0590788	84.633
21	17	.640194	.0716497	66.461	14	20	.509272	.175398	.0563567	84.496
22	18	.622092	.0684545	66.401	14	21	.542190	.175587	.0539094	84.317
23	19	.605709	.0655900	66.288	15	22	.531650	.168092	.0516082	84.247
24	20	.590798	.0630065	66.131	16	23	.521895	.161307	.0495251	84.132
25	20	.652475	.0605006	66.115	17	24	.512834	.155136	.0476304	83.980
26	21	.636502	.0581740	66.115	17	25	.542341	.155498	.0458907	83.811
27	22	.621843	.0560556	66.072	18	26	.533236	.149905	.0442401	83.718
28	23	.608333	.0541184	65.993	19	27	.524718	.144765	.0427233	83.594
29	24	.595834	.0523395	65.883	18	27	.529588	.203622	.0412679	83.558
30	24	.647255	.0505915	65.887	19	28	.521922	.196822	.0398898	83.564
31	25	.634017	.0489616	65.884	20	29	.514693	.190530	.0386145	83.539
32	26	.621699	.0474550	65.852	21	30	.507861	.184689	.0374308	83.487
33	27	.610203	.0460582	65.793	21	31	.529266	.184782	.0363033	83.471
34	28	.599445	.0447593	65.711	22	32	.522441	.179406	.0352472	83.444
35	28	.643532	.0434715	65.724	23	33	.515963	.174384	.0342605	83.395
36	29	.632230	.0422665	65.721	23	34	.536163	.174543	.0333365	83.325
37	30	.621609	.0411405	65.694	24	35	.529688	.169881	.0324462	83.298
38	31	.611604	.0400859	65.649	25	36	.523519	.165502	.0316097	83.252
39	31	.651175	.0390941	65.588	26	37	.517635	.161380	.0308224	83.189
40	32	.640744	.0381083	65.603	26	38	.536491	.161608	.0300780	83.117

TABLE 1 (Continued)

n	From 1 Order Statistic, $x_k$				From 2 Order Statistics, $x_l$ and $x_m$					
	k	$c_k$	$V_k/\sigma^2$	$E_k(\%)$	l	m	$c_l$	$c_m$	$V_{lm}/\sigma^2$	$E_{lm}(\%)$
41	33	.630885	.0371813	65.508	27	39	.530595	.157745	.0293590	83.076
42	34	.621548	.0363080	65.577	27	39	.511693	.194092	.0286687	83.051
43	35	.612692	.0354837	65.539	27	40	.527739	.194098	.0280016	83.052
44	35	.647770	.0346995	65.497	28	41	.522497	.189685	.0273649	83.053
45	36	.638578	.0339230	65.508	29	42	.517464	.185500	.0267612	83.039
46	37	.629834	.0331879	65.503	30	43	.512624	.181525	.0261877	83.013
47	38	.621506	.0324908	65.485	30	44	.527790	.181619	.0256330	83.005
48	39	.613561	.0318289	65.454	31	45	.522953	.177867	.0251035	82.990
49	39	.645063	.0311934	65.425	32	46	.518293	.174292	.0245990	82.963
50	40	.636847	.0305660	65.432	32	47	.532844	.174417	.0241175	82.927
51	41	.628992	.0299688	65.427	33	48	.528185	.171029	.0236490	82.912
52	42	.621475	.0293996	65.412	34	49	.523688	.167791	.0232014	82.886
53	43	.614272	.0288564	65.386	35	50	.519344	.164695	.0227732	82.851
54	43	.642858	.0283310	65.365	35	51	.533183	.164857	.0223621	82.812
55	44	.635431	.0278136	65.370	35	51	.518358	.192720	.0219586	82.800
56	45	.628302	.0273189	65.366	36	52	.514374	.189295	.0215684	82.793
57	46	.621452	.0268453	65.352	36	53	.526631	.189331	.0211879	82.794
58	47	.614863	.0263915	65.329	37	54	.522652	.186069	.0208246	82.793
59	47	.641028	.0259499	65.315	38	55	.518794	.182935	.0204738	82.784
60	48	.634252	.0255159	65.319	39	56	.515050	.179922	.0201366	82.768
61	49	.627725	.0250994	65.314	39	57	.526789	.180005	.0198076	82.763
62	50	.621434	.0246992	65.302	40	58	.523047	.177123	.0194905	82.753
63	51	.615363	.0243145	65.282	41	59	.519411	.174347	.0191850	82.737
64	51	.639486	.0239381	65.273	41	60	.530779	.174447	.0188902	82.715
65	52	.633255	.0235689	65.275	42	61	.527144	.171785	.0186019	82.704
66	53	.627237	.0232133	65.271	43	62	.523609	.169216	.0183238	82.688
67	54	.621420	.0228707	65.260	44	63	.520168	.166736	.0180552	82.665
68	55	.615792	.0225404	65.242	43	63	.522477	.191905	.0177944	82.643
69	55	.638167	.0222158	65.236	44	64	.519185	.189124	.0175365	82.644
70	56	.632401	.0218979	65.238	45	65	.515976	.186434	.0172871	82.638
71	57	.626818	.0215909	65.234	45	66	.525891	.186478	.0170434	82.639
72	58	.621409	.0212943	65.224	46	67	.522686	.183891	.0168070	82.638
73	59	.616163	.0210076	65.208	47	68	.519559	.181386	.0165781	82.631
74	59	.637027	.0207248	65.204	48	69	.516508	.178961	.0163564	82.619
75	60	.631661	.0204481	65.206	48	70	.526081	.179033	.0161388	82.617
76	61	.626455	.0201804	65.201	49	71	.523031	.176693	.0159279	82.609
77	62	.621399	.0199211	65.192	50	72	.520052	.174423	.0157233	82.597
78	63	.616488	.0196699	65.178	50	73	.529377	.174507	.0155244	82.583
79	63	.636031	.0194214	65.177	51	74	.526398	.172314	.0153294	82.575
80	64	.631014	.0191784	65.177	52	75	.523487	.170185	.0151399	82.563

TABLE 1 (concluded)

n	From 1 Order Statistic, $x_k$				From 2 Order Statistics, $x_l$ and $x_m$					
	k	$c_k$	$V_k/\sigma^2$	$E_k$ (%)	l	m	$c_l$	$c_m$	$V_{lm}/\sigma^2$	$E_{lm}$ (%)
81	65	.626136	.0189428	65.173	53	76	.520640	.168116	.0149559	82.547
82	66	.621392	.0187142	65.165	52	76	.522467	.189026	.0147749	82.540
83	67	.616774	.0184924	65.152	53	77	.519723	.186749	.0145969	82.539
84	67	.635154	.0182722	65.152	54	78	.517038	.184536	.0144240	82.535
85	68	.630443	.0180572	65.153	54	79	.525362	.184580	.0142541	82.536
86	69	.625856	.0178483	65.149	55	80	.522678	.182437	.0140886	82.534
87	70	.621385	.0176452	65.141	56	81	.520050	.180352	.0139276	82.529
88	71	.617028	.0174478	65.129	56	82	.528185	.180407	.0137707	82.520
89	71	.634376	.0172514	65.131	57	83	.525558	.178385	.0136163	82.518
90	72	.629936	.0170598	65.130	58	84	.522984	.176415	.0134660	82.512
91	73	.625605	.0168733	65.127	59	85	.520461	.174496	.0133195	82.503
92	74	.621380	.0166917	65.119	59	86	.528366	.174567	.0131763	82.493
93	75	.617256	.0165150	65.109	60	87	.525843	.172703	.0130356	82.487
94	75	.633681	.0163387	65.111	61	88	.523368	.170884	.0128984	82.478
95	76	.629482	.0161668	65.111	62	89	.520941	.169111	.0127645	82.466
96	77	.625382	.0159993	65.107	61	89	.522452	.186991	.0126315	82.466
97	78	.621376	.0158360	65.100	62	90	.520100	.185065	.0125015	82.465
98	78	.637131	.0156767	65.091	63	91	.517792	.183185	.0123745	82.461
99	79	.633057	.0155177	65.093	63	92	.524964	.183228	.0122493	82.462
100	80	.629074	.0153627	65.093	64	93	.522657	.181399	.0121271	82.460

TABLE 2

Best Estimators from 2 Order Statistics  $x_1, x_m$  of Parameters  $\alpha, \sigma$  and Mean  $\mu$  of  
2-Parameter Exponential Population

$n$	$m$	$c_\alpha$	$c_\sigma$	$V_{\hat{\alpha}}/\sigma^2$	$E_{\hat{\alpha}}(\%)$	$V_{\hat{\sigma}}/\sigma^2$	$E_{\hat{\sigma}}(\%)$	$V_{\hat{\mu}}/\sigma^2$	$E_{\hat{\mu}}(\%)$
2	2	.500000	1.00000	.5000000	100.00	1.000000	100.00	.5000000	100.00
3	3	.222222	.666667	.1728395	96.429	.5555556	90.000	.3580247	93.103
4	4	.136364	.545455	.08780992	94.902	.4049587	82.313	.2902893	86.121
5	5	.096000	.480000	.05312000	94.127	.3280000	76.220	.2499200	80.026
6	6	.072993	.437956	.03557580	93.697	.2807289	71.243	.2227284	74.830
7	6	.098522	.689655	.02517789	94.565	.2337165	71.311	.1921182	74.359
8	7	.078475	.627803	.01877684	95.102	.2017178	70.820	.1700652	73.501
9	8	.064680	.582121	.01455215	95.442	.1787246	69.940	.1535601	72.357
10	9	.054676	.546756	.01161360	95.673	.1613595	68.859	.1407012	71.073
11	9	.063619	.699806	.009477724	95.919	.1468046	68.118	.1295906	70.151
12	10	.054829	.657948	.007870456	96.256	.1333457	68.175	.1189919	70.033
13	11	.047981	.623748	.006642280	96.507	.1225454	68.002	.1103346	69.718
14	12	.042514	.595191	.005682027	96.700	.1136773	67.668	.1031197	69.268
15	13	.038061	.570919	.004916701	96.852	.1062578	67.222	.0970068	68.724
16	13	.042091	.673448	.004294816	97.016	.0994728	67.020	.0913336	68.430
17	14	.038015	.646247	.003782806	97.189	.0932310	67.038	.0860455	68.363
18	15	.034588	.622580	.003357619	97.330	.0878686	66.945	.0814630	68.197
19	16	.031672	.601766	.003000580	97.447	.0832093	66.766	.0774510	67.955
20	17	.029165	.583292	.002697803	97.545	.0791212	66.520	.0739069	67.653
21	17	.031444	.660325	.002438181	97.653	.0752377	66.456	.0705104	67.535
22	18	.029100	.640195	.002214152	97.758	.0716497	66.461	.0673502	67.490
23	19	.027047	.622092	.002019763	97.847	.0684545	66.401	.0645217	67.385
24	20	.025238	.605709	.001849983	97.925	.0655900	66.288	.0619741	67.232
25	21	.023632	.590798	.001700810	97.993	.0630065	66.131	.0596668	67.039
26	21	.025095	.652475	.001568788	98.067	.0605006	66.115	.0574155	66.988
27	22	.023574	.636502	.001451542	98.137	.0581740	66.115	.0553164	66.955
28	23	.022209	.621843	.001347010	98.199	.0560557	66.072	.0533987	66.882
29	24	.020977	.608333	.001253411	98.254	.0541184	65.993	.0516395	66.776
30	25	.019861	.595834	.001169266	98.303	.0523395	65.883	.0500195	66.641
31	25	.020879	.647255	.001093227	98.357	.0505915	65.887	.0484208	66.620
32	26	.019813	.634017	.001024377	98.408	.0489616	65.884	.0469259	66.594
33	27	.018839	.621699	.000961850	98.453	.0474551	65.852	.0455408	66.540
34	28	.017947	.610203	.000904895	98.494	.0460582	65.793	.0442538	66.462
35	29	.017127	.599445	.000852865	98.531	.0447593	65.711	.0430545	66.361
36	29	.017876	.643532	.000805148	98.572	.0434715	65.724	.0418616	66.356
37	30	.017087	.632230	.000761334	98.610	.0422665	65.721	.0407431	66.335
38	31	.016358	.621609	.000721011	98.644	.0411405	65.694	.0396962	66.293
39	32	.015682	.611604	.000683817	98.676	.0400859	65.649	.0387140	66.232
40	32	.016279	.651175	.000649434	98.705	.0390941	65.588	.0377889	66.157



TABLE 2 (continued)

$n$	$m$	$C\alpha$	$C\sigma$	$V_{\hat{\alpha}}/\sigma^2$	$E_{\hat{\alpha}}(\%)$	$V_{\hat{\sigma}}/\sigma^2$	$E_{\hat{\sigma}}(\%)$	$V_{\hat{\mu}}/\sigma^2$	$E_{\hat{\mu}}(\%)$
41	33	.015628	.640744	.000617554	98.737	.0381083	65.602	.0368669	66.158
42	34	.015021	.630885	.000587971	98.767	.0371813	65.598	.0359988	66.140
43	35	.014455	.621548	.000560469	98.794	.0363080	65.577	.0351797	66.106
44	36	.013925	.612692	.000534857	98.819	.0354837	65.539	.0344057	66.057
45	36	.014395	.647770	.000510963	98.843	.0346995	65.497	.0336683	66.003
46	37	.013882	.638578	.000488621	98.868	.0339230	65.508	.0329368	66.003
47	38	.013401	.629835	.000467717	98.892	.0331879	65.503	.0322434	65.987
48	39	.012948	.621506	.000448130	98.914	.0324909	65.485	.0315852	65.959
49	40	.012522	.613562	.000429750	98.934	.0318289	65.454	.0309596	65.919
50	40	.012901	.645063	.000412477	98.954	.0311934	65.425	.0303581	65.880
51	41	.012487	.636847	.000396219	98.975	.0305661	65.432	.0297636	65.879
52	42	.012096	.628992	.000380906	98.994	.0299689	65.427	.0291971	65.865
53	43	.011726	.621475	.000366465	99.012	.0293996	65.412	.0286567	65.841
54	44	.011375	.614272	.000352831	99.029	.0288564	65.385	.0281405	65.807
55	44	.011688	.642858	.000339944	99.046	.0283310	65.365	.0276407	65.779
56	45	.011347	.635431	.000327747	99.063	.0278136	65.370	.0271480	65.777
57	46	.011023	.628302	.000316195	99.079	.0273189	65.366	.0266765	65.765
58	47	.010715	.621452	.000305245	99.094	.0268453	65.352	.0262249	65.744
59	48	.010421	.614864	.000294855	99.109	.0263915	65.329	.0257918	65.715
60	48	.010684	.641029	.000284986	99.123	.0259499	65.315	.0253699	65.695
61	49	.010398	.634253	.000275602	99.137	.0255159	65.319	.0249549	65.692
62	50	.010125	.627726	.000266675	99.151	.0250994	65.314	.0245564	65.682
63	51	.009864	.621434	.000258176	99.164	.0246992	65.302	.0241733	65.663
64	52	.009615	.615364	.000250077	99.176	.0243145	65.282	.0238047	65.638
65	52	.009838	.639486	.000242352	99.188	.0239381	65.272	.0234439	65.623
66	53	.009595	.633256	.000234979	99.200	.0235689	65.275	.0230896	65.620
67	54	.009362	.627237	.000227938	99.212	.0232133	65.271	.0227483	65.611
68	55	.009139	.621420	.000221209	99.223	.0228708	65.260	.0224193	65.595
69	56	.008925	.615792	.000214774	99.234	.0225404	65.242	.0221018	65.573
70	56	.009117	.638167	.000208615	99.244	.0222159	65.236	.0217897	65.562
71	57	.008907	.632402	.000202717	99.255	.0218979	65.238	.0214838	65.559
72	58	.008706	.626818	.000197066	99.265	.0215909	65.234	.0211882	65.550
73	59	.008512	.621409	.000191648	99.275	.0212943	65.224	.0209025	65.536
74	60	.008327	.616164	.000186451	99.284	.0210076	65.208	.0206263	65.516
75	60	.008494	.637027	.000181462	99.294	.0207248	65.204	.0203536	65.508
76	61	.008311	.631662	.000176670	99.303	.0204481	65.206	.0200867	65.505
77	62	.008136	.626455	.000172066	99.312	.0201804	65.201	.0198283	65.497
78	63	.007967	.621399	.000167640	99.320	.0199211	65.192	.0195779	65.484
79	64	.007804	.616488	.000163382	99.328	.0196699	65.178	.0193353	65.467
80	64	.007950	.636031	.000159285	99.337	.0194214	65.177	.0190951	65.462

TABLE 2 (concluded)

$n$	$m$	$c_\alpha$	$c_\sigma$	$V_{\bar{\alpha}}/\sigma^2$	$E_{\bar{\alpha}}(\%)$	$V_{\bar{\sigma}}/\sigma^2$	$E_{\bar{\sigma}}(\%)$	$V_{\bar{\mu}}/\sigma^2$	$E_{\bar{\mu}}(\%)$
81	65	.007790	.631015	.000155339	99.345	.0191784	65.177	.0188602	65.459
82	66	.007636	.626137	.000151538	99.353	.0189428	65.173	.0186324	65.451
83	67	.007487	.621392	.000147875	99.360	.0187143	65.165	.0184112	65.440
84	68	.007343	.616774	.000144344	99.367	.0184924	65.152	.0181964	65.424
85	68	.007472	.635155	.000140937	99.375	.0182722	65.152	.0179832	65.420
86	69	.007331	.630444	.000137650	99.382	.0180572	65.152	.0177749	65.418
87	70	.007194	.625856	.000134476	99.389	.0178483	65.149	.0175725	65.411
88	71	.007061	.621386	.000131411	99.396	.0176453	65.141	.0173756	65.400
89	72	.006933	.617029	.000128449	99.402	.0174478	65.129	.0171842	65.385
90	72	.007049	.634377	.000125587	99.409	.0172515	65.130	.0169937	65.384
91	73	.006922	.629936	.000122818	99.415	.0170598	65.130	.0168077	65.381
92	74	.006800	.625606	.000120141	99.421	.0168733	65.127	.0166266	65.374
93	75	.006682	.621380	.000117550	99.427	.0166918	65.119	.0164503	65.364
94	76	.006567	.617256	.000115042	99.433	.0165150	65.109	.0162786	65.351
95	76	.006670	.633682	.000112614	99.439	.0163387	65.111	.0161074	65.351
96	77	.006557	.629483	.000110261	99.445	.0161668	65.111	.0159403	65.348
97	78	.006447	.625382	.000107982	99.451	.0159993	65.107	.0157774	65.342
98	79	.006341	.621376	.000105772	99.456	.0158360	65.100	.0156186	65.333
99	79	.006436	.637131	.000103630	99.461	.0156767	65.091	.0154636	65.321
100	80	.006331	.633057	.000101552	99.467	.0155177	65.093	.0153089	65.321

# ON THE TWO SAMPLE PROBLEM: A HEURISTIC METHOD FOR CONSTRUCTING TESTS<sup>1</sup>

BY V. P. GODAMBE

*Science College, Nagpur, India*

**1. Introduction.** The two-sample problem arises as follows. We are given two independent samples from populations  $A$  and  $B$  respectively and are required to investigate whether the population  $A$  could be considered as identical with  $B$ . In the usual terminology of hypothesis testing: Given two independent samples  $x_1, \dots, x_m$  and  $x_{m+1}, \dots, x_{m+n}$  from populations with unknown cumulative distribution functions  $F$  and  $G$  respectively, the problem is to test the composite hypothesis

$$H_0: F = G$$

against the alternatives

$$H_1: F \neq G,$$

$F$  and  $G$  being completely or partially unspecified.

In the following lines we shall discuss a method (subsequently called the  $V$ -method), for testing  $H_0$  against  $H_1$ , when  $F$  and  $G$  are partially specified (the exact meaning of this will be clear later). A test for the situation where  $F$  and  $G$  are completely unspecified is also put forward.

**2. Notation.** Suppose  $F(x)$  and  $G(x)$  to be two cumulative distribution functions on the real axis,  $-\infty < x < \infty$ , such that their frequency functions exist everywhere. Let  $x_1, \dots, x_m$  and  $x_{m+1}, \dots, x_{m+n}$  denote independent samples from  $F$  and  $G$  respectively. Now the combined sample from  $F$  and  $G$  can be represented as a point

$$(2.1) \quad \mathbf{x} = (x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n})$$

in the  $m + n$  dimensional Euclidean space  $\mathfrak{X}$  of all such points. It follows from the existence of the frequency functions that the probability measure of the set of points  $\mathbf{x}$  in  $\mathfrak{X}$  defined by  $x_i = x_j$  for  $i \neq j$  is zero. Next we define on  $\mathfrak{X}$  a vector-valued function  $\gamma$ ,

$$(2.2) \quad \gamma(\mathbf{x}) = (\gamma_1(\mathbf{x}), \dots, \gamma_i(\mathbf{x}), \dots, \gamma_{m+n}(\mathbf{x})),$$

where  $\gamma_i(\mathbf{x})$  is the total number of the components of  $\mathbf{x}$  less than or equal to  $x_i$ . Thus  $\gamma_i(\mathbf{x})$  is the rank of  $x_i$  in the combined sample  $\mathbf{x} = (x_1, \dots, x_{m+n})$ . Further we arrange the last  $n$  components of  $\mathbf{x}$ , that is  $x_{m+1}, \dots, x_{m+n}$ , accord-

Received July 24, 1957; revised June 9, 1959.

<sup>1</sup> This work was submitted as a part of a Ph.D. thesis of the London University. The paper was supported (in part) by funds provided under contract AF-18 (600) 456 with the School of Aviation Medicine, USAF, Randolph Air Force Base, Texas.

ing to their magnitudes as  $-\infty < y_1 < \dots < y_n < \infty$  to define another vector-valued function  $\mathbf{a}$ ,

$$(2.3) \quad \mathbf{a}(\mathbf{x}) = (a_1(\mathbf{x}), \dots, a_i(\mathbf{x}), \dots, a_{n+1}(\mathbf{x})),$$

where  $a_i(\mathbf{x})$  is the total number of the first  $m$  components of  $\mathbf{x}$  lying between  $y_{i-1}$  and  $y_i$ ,  $y_0$  denoting  $-\infty$  and  $y_{n+1}$  denoting  $+\infty$  for convenience. In addition we define

$$(2.4) \quad \mathbf{b}(\mathbf{x}) = (b_1(\mathbf{x}), \dots, b_i(\mathbf{x}), \dots, b_{m+1}(\mathbf{x})),$$

where  $b_i(\mathbf{x})$  denotes the number of individuals out of  $x_{m+1}, \dots, x_{m+n}$  lying between the  $i-1$ st and  $i$ th ordered individuals from  $x_1, \dots, x_m$ ;  $b_1(\mathbf{x})$  and  $b_{m+1}(\mathbf{x})$  being defined analogously to  $a_1(\mathbf{x})$  and  $a_{n+1}(\mathbf{x})$  in (2.3). Now it is important to note that, given  $\mathbf{a}(\mathbf{x})$  in (2.3),  $\mathbf{b}(\mathbf{x})$  in (2.4) is uniquely determined and conversely.

For simplicity we write  $\gamma$  for  $\gamma(\mathbf{x})$ ,  $\mathbf{a}$  for  $\mathbf{a}(\mathbf{x})$ , etc. Now  $P(\gamma | F, G)$  denotes the probability of obtaining  $\mathbf{x}$  such that  $\gamma(\mathbf{x}) = \gamma$  given  $F$  and  $G$ . Similarly, we have  $P(\mathbf{a} | F, G)$ , etc.

**3. The most powerful rank test.** Following the above notation it is easy to see that

$$P(\gamma | F, F) = 1/(m+n)!$$

Hence the most powerful rank test of the hypothesis  $H_0: F = G$ , against the simple alternative  $H_1$ , that the c.d.f.'s are specifically  $F$  and  $G$  respectively, has the critical region

$$(3.2) \quad \gamma: P(\gamma | F, G) > \text{const.}$$

Since hereafter there is no possibility of confusion, we shall write  $P(\gamma)$  for  $P(\gamma | F, G)$ ,  $P(\mathbf{a})$  for  $P(\mathbf{a} | F, G)$ , etc. Now from the definition of

$$\mathbf{a} = (a_1, \dots, a_{n+1})$$

in Section 2 it follows that

$$(3.3) \quad P(\mathbf{a}) = m! n! P(\gamma),$$

which of course is also true under the null hypothesis. Thus the most powerful test (3.2) is associated with the critical region

$$(3.4) \quad \mathbf{a}: P(\mathbf{a}) > \text{const.}$$

Suppose further that we have a function  $\theta$  such that

$$(3.5) \quad G(x) = \theta(F(x))^2$$

<sup>2</sup> Here one should avoid the mistake of assuming that  $F$  is a uniform distribution on  $(0, 1)$ .

for every  $x$  and

$$(3.6) \quad \theta'(F) = (\partial/\partial F)\theta(F)$$

exists for every  $F$ ,  $0 \leq F \leq 1$ .

THEOREM 3. We have

$$(3.7) \quad P(\mathbf{a}) = \frac{m!}{\prod_{i=1}^{n+1} a_i!} \int \cdots \int_D \prod_{i=1}^{n+1} p_i^{a_i} \cdot \prod_{i=1}^n \theta'(p_1 + \cdots + p_i) \prod_{i=1}^n dp_i,$$

where  $p_{n+1} = 1 - p_1 - \cdots - p_n$  and the domain of integration  $D$  is

$$D = \{p_1, \cdots, p_n : 0 \leq p_i \leq 1, \sum p_i \leq 1, i = 1, \cdots, n\}.$$

The theorem can be easily derived from a result of Hoeffding's [1], top of p. 88. It is important to note that, in the above formulae,  $P(\mathbf{a})$  depends on  $F$  and  $G$  only through  $\theta$ , or rather  $\theta'$ , and so does the corresponding test (3.4) for testing against the alternative  $G = \theta(F)$ . It is, however, seldom possible to evaluate the integral on the right side of (3.7). For this and other reasons we shall in the next section put forward another rank test which depends on  $\theta'$  alone.

**4. The V-test.** Consider the following degenerate case of testing a *simple* hypothesis  $H_0$  against a *simple* alternative  $H_1$ . (Note that these are not the same as the hypotheses in Section 1.)

$H_0$ : Both samples  $x_1, \cdots, x_m$  and  $y_1, \cdots, y_n$  are drawn from a common specified c.d.f.  $F$ .

$H_1$ : The sample  $x_1, \cdots, x_m$  is drawn from  $F$ , while the sample  $y_1, \cdots, y_n$  is drawn from another specified c.d.f.  $G$ .

The most powerful test (not the most powerful rank test) in this case is independent of the sample from  $F$  and in fact is given by the critical region

$$(4.1) \quad y_1, \cdots, y_n : \frac{g(y_1) \cdots g(y_n)}{f(y_1) \cdots f(y_n)} > \text{const.},$$

where  $g$  and  $f$  are the frequency functions of  $G$  and  $F$  respectively. Again, if, as in (3.5), we have  $G(x) = \theta(F(x))$  for all  $x$  and if  $\theta'(F)$  exists for  $0 \leq F \leq 1$ , then the critical region (4.1) can be expressed as

$$(4.2) \quad F(y_1), \cdots, F(y_n) : \prod_{i=1}^n \theta'(F(y_i)) > \text{const.}$$

For instance, when  $F$  and  $G$  are normal distributions with unit variance, the mean of  $F$  being 0 and that of  $G$  being  $\delta$ , it can be seen from equation (6.4) of Section 6 that (4.2) can be written as

$$F(y_1), \cdots, F(y_n) : \sum_{i=1}^n \psi^{-1}(F(y_i)) > \text{const.},$$

where  $\psi^{-1}$  is the functional inverse of the normal integral as defined in (6.1). Since, however,  $F$  itself is the standard normal c.d.f., the above critical region is identical with

$$y_1, \dots, y_n: \sum_{i=1}^n y_i > \text{const.}$$

This is the usual optimum test for the normal mean when the variance is known.

Now the critical region (4.2) clearly depends on  $F$  in addition to  $\theta'$ . However, an approximation to (4.2) which depends on  $\theta'$  alone can be worked out as follows. For a given second sample  $(y_1, \dots, y_n)$  the quantities

$$(4.3) \quad \left| F(y_i) - \frac{a_1 + \dots + a_i}{m} \right|$$

can simultaneously be made arbitrarily small for  $i = 1, \dots, n$ , with as large a probability as we please, by increasing sufficiently the size of the first sample. Hence

$$(4.4) \quad a_1, \dots, a_n: \prod_{i=1}^n \theta' \left( \frac{a_1 + \dots + a_i}{m} \right) > \text{const.}$$

is the suggested approximation to (4.2). Note that (4.4) is a non-parametric test, depending only on the order relationships within the sample.

Now for testing the null hypothesis  $H_0: G = F$  against the alternative  $H_1: G = \theta(F)$ , we propose the  $V$ -statistic

$$(4.5) \quad V(\mathbf{a}) = \prod_{i=1}^n \theta' \left( \frac{1 + a_1 + \dots + a_i}{m + 2} \right)$$

or a suitable monotonic increasing function of the right side of (4.5), the corresponding  $V$ -test being defined by the critical region

$$(4.6) \quad \mathbf{a}: V(\mathbf{a}) > \text{const.}$$

This will also be referred to as the  $V$ -method of obtaining tests. The motivation for the  $V$ -method is made clear in the preceding paragraph. In fact, (4.6) is obtained from (4.4), with a small modification to prevent the  $V$ -statistic from assuming infinitely large values.

About the intuitive appeal of the  $V$ -method, it may be said that some tests derived by its application, with a slight difference, have already been proposed by different authors on more or less intuitive grounds. This will be verified in some of the subsequent sections, where  $V$ -tests are compared with some of the known tests, including the one given by the statistic (3.7).

Further it would appear from the above discussion that, though the  $V$ -method is put forward as a sure method of obtaining tests for simple alternatives, it can in some cases yield tests even for composite alternatives. This can be checked from the illustrations to follow.



5. **Lehmann's alternatives.** In this case it is assumed that

$$(5.1) \quad G = \theta(F) = F^k,$$

where  $k > 1$ . From (5.1) we have

$$(5.2) \quad \theta'(F) = kF^{k-1}.$$

Hence from (4.6) the  $V$ -test for the present situation is given by the critical region

$$(5.3) \quad \mathbf{a}: V(\mathbf{a}) > \text{const.},$$

where the  $V$ -statistic is defined by

$$(5.4) \quad V(\mathbf{a}) = \prod_{i=1}^n (1 + a_1 + \cdots + a_i).$$

It is interesting to see that the  $V$ -test does not depend upon  $k$  for  $k > 1$ . The test for  $k < 1$  can be obtained similarly.

I. R. Savage [3] has studied very extensively the alternatives in (5.1). He also has tabulated the probabilities  $P(\mathbf{a})$  in (3.7), when  $\theta(F) = F^k$ , for different values of  $k$  and different sample sizes.

In Table 1 we give the 5  $\mathbf{a}$ 's corresponding to the largest values of  $V(\mathbf{a})$  in (5.4). Now it so happens that the same  $\mathbf{a}$ 's are the ones having the largest probabilities  $P(\mathbf{a})$  in (3.7), for all values of  $k > 1$  considered by Savage [3]. For these values of  $k$ , the ordering of the  $P(\mathbf{a})$ 's mentioned above is also the same. Hence in Table 1 the  $P(\mathbf{a})$ 's are reproduced from Savage's table for just one set of the values of  $k$ . From them we can construct the most powerful tests, defined by the critical regions  $\mathbf{a}: P(\mathbf{a}) > \text{const.}$ , as in (3.4), up to the significance level of 25 per cent. Each of these tests, as can be seen from Table 1, will be equal in power to the corresponding  $V$ -test. Further from Savage's table in [3] it appears that the performance of the  $V$ -test for sample sizes other than those considered in Table 1 is equally good.

The statistic that Savage [3] has proposed for the present problem is, in our notation,

$$(5.5) \quad T(\mathbf{a}) = \sum_{i=1}^{n+1} \sum_{j=0}^{a_i} \frac{a_1 + \cdots + a_{i-1} + j}{a_1 + \cdots + a_{i-1} + j + (i-1)},$$

the corresponding critical region being

$$(5.6) \quad \mathbf{a}: T(\mathbf{a}) < \text{const.}$$

The 5 smallest values of  $T(\mathbf{a})$  are reproduced in Table 1. It is difficult to see any connection between (5.4) and (5.6). Now Savage [3] has proved that for the cases  $m = 2$ ,  $n = 3$  and  $m = 2$ ,  $n = 4$ , dealt with in Table 1, the simple ordering of the probabilities  $P(\mathbf{a})$  in (3.7) when  $\theta(F) = F^k$  does not depend on  $k$  for  $k > 1$ , and is given by the statistic  $T(\mathbf{a})$  in (5.5).

TABLE 1

$\mathbf{a}$	$P(\mathbf{a})$ for $G = F^k$ $k = 3.7769$	$V(\mathbf{a})$ as in (5.4)	$T(\mathbf{a})$ as in (5.5)
$m = 2 \quad n = 3$			
(2, 0, 0, 0)	.4394	8	1.4333
(1, 1, 0, 0)	.1840	4	1.9333
(1, 0, 1, 0)	.1242	2	2.2667
(1, 0, 0, 1)	.0963	1	2.5167
(0, 2, 0, 0)	.0487	0	2.9333
$m = 2 \quad n = 4$			
$\mathbf{a}$	$P(\mathbf{a})$ for $G = F^k$ $k = 3.6173$	$V(\mathbf{a})$ as in (5.4)	$T(\mathbf{a})$ as in (5.5)
(2, 0, 0, 0, 0)	.3743	16	2.1000
(1, 1, 0, 0, 0)	.1621	8	2.6000
(1, 0, 1, 0, 0)	.1106	4	2.9333
(1, 0, 0, 1, 0)	.0862	2	3.1833
(1, 0, 0, 0, 1)	.0716	1	3.3833
$m = 3 \quad n = 3$			
$\mathbf{a}$	$P(\mathbf{a})$ for $G = F^k$ $k = 3.0546$	$V(\mathbf{a})$ as in (5.4)	$T(\mathbf{a})$ as in (5.5)
(3, 0, 0, 0)	.2549	27	1.1500
(2, 1, 0, 0)	.1513	18	1.4833
(2, 0, 1, 0)	.1130	12	1.7333
(2, 0, 0, 1)	.0922	8	1.9333
(1, 2, 0, 0)	.0746	9	1.9833

Next we consider the alternative

$$(5.7) \quad G = \theta(F) = \lambda F^2 + (1 - \lambda)F, \quad 0 < \lambda < 1.$$

As proved by Lehmann [2], the well-known Wilcoxon test is optimum against the alternative hypothesis (5.7) for very small values of  $\lambda$ . It is interesting to see that the  $V$ -statistic (4.5) for the present situation is again Wilcoxon's statistic. For from (5.7) we have

$$(5.8) \quad \theta'(F) = 2\lambda F + (1 - \lambda).$$

Now for very small values of  $\lambda$ , (5.8) can be written as

$$(5.9) \quad \theta'(F) = e^{2\lambda F}.$$

Hence the  $V$ -statistic is

$$(5.10) \quad V(\mathbf{a}) = \sum_{i=1}^n \frac{1 + a_i + \cdots + a_i}{m+2},$$

the corresponding  $V$ -test being defined by the critical region

$$(5.11) \quad \mathbf{a}: V(\mathbf{a}) > \text{const.}$$

This is the usual one-sided Wilcoxon test.

**6. Normal alternatives.** Let  $F$  and  $G$  be the following distributions:

$$(6.1) \quad F(x) = (2\pi)^{-1} \int_{-\infty}^x \exp(-\frac{1}{2}h^2) dh = \psi(x)$$

and

$$(6.2) \quad \begin{aligned} G(x) &= (2\pi)^{-1} \int_{-\infty}^x \exp(-\frac{1}{2}(h+\delta)^2) dh \\ &= (2\pi)^{-1} \int_{-\infty}^{x+\delta} \exp(-\frac{1}{2}h^2) dh = \psi(x+\delta). \end{aligned}$$

Now though for convenience of notation it has been assumed above that both  $F$  and  $G$  have variances equal to 1, the following arguments are valid for any unknown common variance  $\sigma^2$ . Write

$$(6.3) \quad G(x) = \theta(F(x)).$$

Then it is seen from (6.1) and (6.2) that

$$(6.4) \quad \begin{aligned} \theta'(F) &= \left( \frac{\partial G}{\partial x} \right) / \left( \frac{\partial F}{\partial x} \right) = \exp(-\delta^2 - 2\delta x) \\ &= \exp(-\delta^2 - 2\delta\psi^{-1}(F)). \end{aligned}$$

Next since  $\theta'(F)$  in (6.4) depends on  $\delta$  it follows from (3.7) that the most powerful rank order test of  $H_0$  against  $H_1$  may in general depend on  $\delta$ , though it has been proved to be independent of  $\delta$  for all sufficiently small values of  $\delta$ . In fact, it is then Hoeffding's  $C_1$  criterion [1]. Furthermore, some empirical sampling investigations by Teichroew [4] suggest that the most powerful rank order test may exist uniquely for all  $\delta > 0$ . The situation for  $\delta < 0$  is similar. However, no theoretical result is available in that direction.

It is interesting to see that in the present case the  $V$ -test obtained from (6.4) above does not depend upon  $\delta$ . It follows from (6.4) and (4.6) that the  $V$ -test is defined by the critical region

$$(6.5) \quad \mathbf{a}: V(\mathbf{a}) > \text{const.},$$

where the  $V$ -statistic is defined by

$$(6.6) \quad V(\mathbf{a}) = -\sum_{i=1}^n \psi^{-1} \left( \frac{1 + a_i + \cdots + a_i}{m+2} \right).$$

In the following illustrations the relative frequencies are reproduced from Teichroew's experiments [4] for some specified  $\delta$ . However, the ordering of the  $\mathbf{a}$ 's by their relative frequencies is more or less the same for the other values of  $\delta$  considered by Teichroew. It can be seen from Table 2 that the ordering of  $\mathbf{a}$  by  $V(\mathbf{a})$  in (6.6) is nearly the same as that by the relative frequencies. The performance of  $V(\mathbf{a})$  is as good for the other sample sizes of Teichroew [4] as for those considered in Table 2. It must, however, be said that for all these illus-

TABLE 2

 $m = 3 \quad n = 2$ 

$\mathbf{a}$	$V(\mathbf{a})$ as in (6.6)	$\sigma = 0.75^a$ Relative Frequency	$X(\mathbf{a})$ as in (6.7)	Hoeffding's $C_1$
(3, 0, 0)	-1.68	2.25	-1.40	-1.66
(2, 1, 0)	-1.09	3.45	-0.97	-1.16
(1, 2, 0)	-0.59	4.45	-0.54	-0.66
(2, 0, 1)	-0.50	5.45	-0.45	-0.50
(1, 1, 1)	-0.00	7.40	-0.00	-0.00
(0, 3, 0)	0.00	8.10	0.00	0.00
(1, 0, 2)	0.50	11.15	0.45	0.50
(0, 2, 1)	0.59	12.00	0.54	0.66
(0, 1, 2)	1.09	18.45	0.97	1.16
(0, 0, 3)	1.68	27.30	1.40	1.66

trative cases Hoeffding's  $C_1$ -test [1] or van der Waerden's  $X$ -test [5] has comparably good performance.

Now van der Waerden's  $X$ -statistic, in our notation is defined as follows.

$$(6.7) \quad X(\mathbf{a}) = -\sum_{i=1}^n \psi^{-1} \left( \frac{a_1 + \cdots + a_i + i}{m + n + 1} \right).$$

When  $m$  is large enough compared to  $n$ ,  $X(\mathbf{a})$  is nearly equal to  $V(\mathbf{a})$  in (6.6). For the case considered in Table 2, we see that even for  $m = 3$  and  $n = 2$ , the critical regions given by the two statistics are identical. It may also be of interest to note that the  $X$ -test has been shown to be asymptotically equivalent to the  $C_1$ -test.

Next we consider two normal populations with the same mean but different variances,

$$(6.8) \quad \begin{aligned} F(x) &= (2\pi\sigma_1^2)^{-1} \int_{-\infty}^x \exp(-\frac{1}{2}h^2/\sigma_1^2) dh \\ &= (2\pi)^{-1} \int_{-\infty}^{x/\sigma_1} \exp(-\frac{1}{2}h^2) dh = \psi(x/\sigma_1) \end{aligned}$$

<sup>a</sup> These are Monte Carlo results; see [4].

and

$$\begin{aligned}
 (6.9) \quad G(x) &= (2\pi\sigma_2^2)^{-1} \int_{-\infty}^x \exp(-\frac{1}{2}h^2/\sigma_2^2) dh \\
 &= (2\pi)^{-1} \int_{-\infty}^{x/\sigma_2} \exp(-\frac{1}{2}h^2) dh = \psi(x/\sigma_2). \quad \}
 \end{aligned}$$

Consider the problem of testing  $H_0: \sigma_1^2 = \sigma_2^2$  against  $\sigma_1^2 < \sigma_2^2$ .

If we put  $G(x) = \theta(F(x))$  from (6.8) and (6.9) we have

$$\begin{aligned}
 (6.10) \quad \theta'(F) &= \left( \frac{\partial G}{\partial x} \right) / \left( \frac{\partial F}{\partial x} \right) = \exp \frac{1}{2} (x^2/\sigma_1^2 - x^2/\sigma_2^2) \\
 &= \exp \left\{ \frac{1}{2} (\psi^{-1}(F))^2 - \frac{1}{2} (\sigma_1^2/\sigma_2^2) (\psi^{-1}(F))^2 \right\} \\
 &= \exp \frac{1}{2} (1 - k^2) (\psi^{-1}(F))^2
 \end{aligned}$$

where  $\sigma_1^2/\sigma_2^2 = k^2$ . Hence, to test  $\sigma_1^2 < \sigma_2^2$ , the critical region of the  $V$ -test would be, from (6.10) and (4.6),

$$(6.11) \quad \mathbf{a}: V(\mathbf{a}) > \text{const.}$$

Here the  $V$ -statistic is given by

$$(6.12) \quad V(\mathbf{a}) = \sum_1^n \left( \psi^{-1} \left( \frac{1 + a_1 + \dots + a_i}{m+2} \right) \right)^2.$$

Unfortunately we do not have the necessary Monte Carlo frequencies to judge empirically for small samples the performance of  $V(\mathbf{a})$  in (6.12). Later on, we shall have an interesting comparison of  $V(\mathbf{a})$  in (6.12) with some other statistics. Furthermore a different application of the  $V$ -method, suggested by the theorem in the next section, gives another test for the same problem of testing the variance ratio of two normal populations with the same mean. This test is discussed at the end of the next section.

**7. A theorem about the  $V$ -statistic.** Consider a case where  $G = \theta(F)$  and

$$(7.1) \quad \frac{\partial^2 \theta(F)}{\partial F^2} \geq 0$$

for all  $F$ , which implies that

$$(7.2) \quad \frac{d}{dx} \frac{\partial G / \partial x}{\partial F / \partial x} \geq 0.$$

Now (7.2) is precisely the monotone likelihood ratio condition of Savage [3]. Thus it follows from Theorem 6.1 of Savage [3] that if

$$\begin{aligned}
 (7.3) \quad \mathbf{a} &= (\dots, a_i, \dots, a_j, \dots) \\
 \mathbf{a}' &= (\dots, a_i + 1, \dots, a_j - 1, \dots),
 \end{aligned}$$

the other components of  $\mathbf{a}$  and  $\mathbf{a}'$  being identical and  $i < j$ , then

$$(7.4) \quad P(\mathbf{a}') \geq P(\mathbf{a}).$$

Further, from (4.5) and (7.1) we have

$$(7.5) \quad V(\mathbf{a}') \geq V(\mathbf{a}).$$

Now the  $V$ -statistic has been defined for a simple alternative  $G = \theta(F)$ , and as such we can order simply the vectors  $\mathbf{a}$ , according to the values  $V(\mathbf{a})$ . Similarly the vectors  $\mathbf{a}$  can be ordered simply according to the values of  $P(\mathbf{a})$  in (3.7). Let  $\mathbf{S}$  be a set of vectors which could be arranged in such a way that for any pair of successive vectors,  $\mathbf{a}_k$  and  $\mathbf{a}_{k+1}$  say,  $\mathbf{a}_k$  is related to  $\mathbf{a}_{k+1}$  as  $\mathbf{a}$  to  $\mathbf{a}'$  in (7.3) for some  $i, j (i < j)$ , where  $i, j$  can vary with  $k$ . Then from (7.3), (7.4) and (7.5) we have

**THEOREM 7.<sup>4</sup>** *For any function  $\theta$  satisfying (7.1), a simple ordering of  $\mathbf{S}$ , according to  $P(\mathbf{a})$  is identical with a simple ordering given by  $V(\mathbf{a})$ .*

There is a similar theorem if instead of (7.1) we have

$$\frac{\partial^2 \theta(F)}{\partial F^2} \leq 0.$$

Now as already noted by Savage [3], for both Lehmann's alternative in (5.1) and the normal alternative in (6.1)–(6.2), the monotone likelihood ratio condition (7.1) is fulfilled. This can also be checked from (5.2) and (6.4). Hence Theorem 7 above is applicable in both cases.

On the other hand, for the alternative in (6.8)–(6.9), the condition (7.1) is not fulfilled. This can be seen from (6.10). We therefore substitute

$$(7.6) \quad \frac{1}{2}x^2 = z.$$

(This transformation is due to the referee.) Now the c.d.f.'s of  $z$ , viz.  $F(z)$  and  $G(z)$ , corresponding to (6.8) and (6.9), are

$$(7.7) \quad F(z) = (\pi)^{-1} \int_0^{z/\sigma_1^2} e^{-h} h^{-1} dh$$

and

$$(7.8) \quad G(z) = (\pi)^{-1} \int_0^{z/\sigma_2^2} e^{-h} h^{-1} dh.$$

Now putting  $G = \theta(F)$ , we have

$$(7.9) \quad \begin{aligned} \theta'(F) &= \text{const. exp } [(1 - \sigma_1^2/\sigma_2^2)z/\sigma_1^2] \\ &= \text{const. exp } [(1 - \sigma_1^2/\sigma_2^2)\sqrt{2}\Gamma^{-1}(F, -\frac{1}{2})] \end{aligned}$$

<sup>4</sup> The author is indebted to the referee for an important clarification in this theorem.



where  $I^{-1}$  is the inverse of  $I$  in (7.14). It is clear from (7.7), (7.8) and (7.9) that the monotone likelihood ratio condition of (7.1) is now satisfied. Next referring to the notation of Section 2, it can be seen that a point

$$\mathbf{x} = (x_1, \dots, x_{m+n})$$

in  $\mathfrak{X}$  is transformed by the substitution (7.6), i.e.,  $\frac{1}{2}x_i^2 = z_i$ ,  $i = 1, \dots, m+n$ , into a point

$$(7.10) \quad \mathbf{z} = (z_1, \dots, z_{m+n})$$

in  $\mathfrak{Z}$ , say. Further, exactly analogous to  $\mathbf{a} = (a_1(\mathbf{x}), \dots, a_{n+1}(\mathbf{x}))$  in (2.3), we can define on  $\mathfrak{Z}$  a vector-valued function

$$(7.11) \quad \mathbf{c}(\mathbf{z}) = (c_1(\mathbf{z}), \dots, c_{n+1}(\mathbf{z})).$$

Now from (4.5), (7.9), and (7.11) the  $V$ -test for testing the null hypothesis against the alternative in (7.7)–(7.8) is defined by the critical region

$$(7.12) \quad \mathbf{c}: V(\mathbf{c}) > \text{const.}$$

where the  $V$ -statistic is given by

$$(7.13) \quad V(\mathbf{c}) = \sum_{i=1}^{n+1} I^{-1} \left( \frac{c_1 + \dots + c_i + 1}{m+2}, -\frac{1}{2} \right),$$

$I^{-1}$  as before being the inverse of

$$(7.14) \quad I(u, -\frac{1}{2}) = (\pi)^{-1} \int_0^{\sqrt{1-u}} e^{-h} h^{-1} dh.$$

(The values of (7.14) for different  $u$ 's are tabulated in *Tables of the Incomplete  $\Gamma$ -Function*, Cambridge University Press, 1946.)

Two suggestions for comparing the statistics (7.13) and (6.12) are as follows: (i) It will be remembered from (6.8) and (6.9) that for convenience of notation we assumed the common mean of the two populations to be zero. Actually, if  $\mu$  is the common mean, the transformation (7.6) would be  $\frac{1}{2}(x - \mu)^2 = z$ , which means the vector  $\mathbf{c}$  in (7.11) depends on  $\mu$ . That is, contrary to the test (6.11), the test (7.12) cannot be worked out unless  $\mu$  is known. (ii) Theorem 7 above is valid for (7.13) while it is not valid for (6.12).

Now Theorem 7 implies some justification for using any of the  $V$ -statistics, such as (5.4) or (6.6), where  $\theta$  satisfies (7.1), for testing against a wider alternative hypothesis  $G = \theta(F)$ , which does not specify anything about  $\theta$ , excepting that it satisfies the monotone likelihood ratio condition (7.1).

In the next section we develop a test of the null hypothesis  $H_0: F = G$  against the general alternative  $H_1: F \neq G$ .

#### 8. $\phi$ -test. Substituting in (3.7)

$$(8.1) \quad \phi(\mathbf{a}/\mathbf{p}) = \frac{m!}{\prod_{i=1}^{n+1} a_i!} \prod_{i=1}^{n+1} (p_i)^{a_i},$$

we have

$$(8.2) \quad P(\mathbf{a}) = \int \cdots \int_D \phi(\mathbf{a}/\mathbf{p}) n! \prod_1^n \theta'(p_1 + \cdots + p_i) \prod_1^n dp_i.$$

Now in (8.1),  $a_i/m$  is the maximum likelihood estimate of  $p_i$ ,  $i = 1, \dots, n$ . Therefore if

$$(8.3) \quad \phi(\mathbf{a}) = \frac{m!}{\prod_1^{n+1} a_i!} \prod_1^{n+1} (a_i/m)^{a_i},$$

it follows that

$$(8.4) \quad \phi(\mathbf{a}) \geq \phi(\mathbf{a}/\mathbf{p}).$$

Thus from (8.1), (8.2) and (8.4) we have

$$(8.5) \quad P(\mathbf{a}) < \phi(\mathbf{a}) \int \cdots \int_D n! \prod_1^n \theta'(p_1 + \cdots + p_i) \prod_1^n dp_i.$$

Now using the transformation  $p_1 + \cdots + p_i = q_i$ ,  $i = 1, \dots, n$  we have

$$(8.6) \quad \int \cdots \int_D \theta'(p_1 + \cdots + p_i) \prod_1^n dp_i = \int \cdots \int_{D'} \prod_1^n \theta'(q_i) \prod_1^n dq_i,$$

where

$$D' = \{q_1, \dots, q_n : 0 \leq q_i \leq 1, q_{i-1} \leq q_i, i = 1, \dots, n\},$$

$q_0$  denoting 0. Integrating the right hand side of (8.6) term by term and noting that  $\theta(0) = 0$  and  $\theta(1) = 1$ , we have

$$(8.7) \quad \int \cdots \int_{D'} \prod_1^n \theta'(q_i) \prod_1^n dq_i = 1/n!.$$

Hence from (8.5), (8.6) and (8.7) it follows that

$$(8.8) \quad P(\mathbf{a}) \leq \phi(\mathbf{a}).$$

Next we recollect the definition of

$$(8.9) \quad \mathbf{b} = (b_1, \dots, b_{m+1})$$

in (2.4) of Section 2. It has also been noted that  $\mathbf{a}$  defines  $\mathbf{b}$  uniquely and conversely. Therefore we have

$$(8.10) \quad P(\mathbf{a}) = P(\mathbf{b}).$$

Further it follows from (8.3) and (8.8) that if

$$(8.11) \quad \phi(\mathbf{b}) = \frac{n!}{\prod_1^{m+1} b_i!} \prod_1^{m+1} (b_i/n)^{b_i},$$

then

$$(8.12) \quad P(\mathbf{b}) \leq \phi(\mathbf{b}).$$

We can write (8.8), (8.10) and (8.12) together as

$$(8.13) \quad P(\mathbf{a}) = P(\mathbf{b}) \leq \min(\phi(\mathbf{a}), \phi(\mathbf{b})).$$

Now define the  $\phi$ -statistic as the minimum of  $\phi(\mathbf{a})$  and  $\phi(\mathbf{b})$  i.e.,

$$(8.14) \quad \phi = \min(\phi(\mathbf{a}), \phi(\mathbf{b})) = \phi(\mathbf{a}, \mathbf{b}).$$

Then the  $\phi$ -test for testing  $H_0: F = G$  against  $H_1: F \neq G$  is defined by the critical region

$$(8.15) \quad \phi > \text{const.}$$

The motivation for this test lies in the inequality (8.13) and the fact that in a degenerate case when there exist two numbers  $u$  and  $v$ ,  $v > u$ , such that  $F(u) = 1$  and  $G(v) = 0$ , then

$$(8.16) \quad P(\mathbf{a}) = P(\mathbf{b}) = \phi = 1$$

with probability equal to unity.

Wolfowitz [6] proposed a test statistic equivalent to

$$(8.17) \quad W = \phi(\mathbf{a}) \cdot \phi(\mathbf{b})$$

for testing  $H_0: F = G$  against  $F \neq G$ . In the numerical illustrations in section 10, the  $\phi$ -statistic defined in (8.14) appears to be better than Wolfowitz' statistic in (8.17), though possibly quite a few statements made hereafter in case of the  $\phi$ -statistic may also hold for  $W$  in (8.17). A simple method for computing  $\phi$ -statistic could be obtained from one suggested by Wolfowitz [6] for his statistic  $W$  above.

#### 9. Some properties of the $\phi$ -statistic. Let

$$(9.1) \quad \mathbf{a} = (\dots, a_i, \dots, a_j, \dots)$$

$$(9.2) \quad \mathbf{a}' = (\dots, a_i + 1, \dots, a_j - 1, \dots)$$

be two vectors with their  $i$ th and  $j$ th components ( $i < j$ ) as shown ( $a_i, a_j \geq 1$ ), other components of  $\mathbf{a}$  being equal to the corresponding ones of  $\mathbf{a}'$ . Now from (8.3) we have

$$(9.3) \quad \frac{\phi(\mathbf{a}')}{\phi(\mathbf{a})} = \frac{(1 + 1/a_i)^{a_i}}{(1 + 1/(a_j - 1))^{a_j - 1}}.$$

Thus from the monotonicity of function  $(1 + 1/z)^z$ , if in (9.1) and (9.2)

$$(9.4) \quad a_i \geq a_j,$$

we have

$$(9.5) \quad \phi(\mathbf{a}') > \phi(\mathbf{a}).$$

Further, it follows from the above argument and the symmetry of  $\phi(\mathbf{a})$  in

$a_1, \dots, a_{n+1}$  that if

$$(9.6) \quad \mathbf{a}'' = (\dots, a_i + a_j, \dots, 0, \dots)$$

with its  $i$ th and  $j$ th components as shown, other components being equal to the corresponding ones of  $\mathbf{a}$  in (9.1), then

$$(9.7) \quad \phi(\mathbf{a}'') > \phi(\mathbf{a})$$

regardless of the condition (9.4) above. Again as in (8.9) define  $\mathbf{b}$  and  $\mathbf{b}'$  from  $\mathbf{a}$  and  $\mathbf{a}'$  in (9.1) and (9.2). Then it follows from arguments similar to the above that

$$(9.8) \quad \phi(\mathbf{b}') \geq \phi(\mathbf{b}).$$

Thus from (9.5) and (9.8) we have for  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{a}'$ ,  $\mathbf{b}'$  defined by (9.1) and (9.2) above, provided (9.4) holds,

$$(9.9) \quad \phi(\mathbf{a}', \mathbf{b}') \geq \phi(\mathbf{a}, \mathbf{b}),$$

$\phi$  being given by (8.14). Similarly from (9.6) and (9.7) we have, regardless of (9.4),

$$(9.10) \quad \phi(\mathbf{a}'', \mathbf{b}'') \geq \phi(\mathbf{a}, \mathbf{b}).$$

Now suppose  $F$  and  $G$  are such  $G(x) = \theta(F(x))$  and

$$(9.11) \quad \frac{\partial^2 \theta(F)}{\partial F^2} \geq 0$$

for all  $F$ . Then as said in Section 7, (9.11) implies

$$(9.12) \quad \frac{d}{dx} \frac{\partial G / \partial x}{\partial F / \partial x} \geq 0$$

for all  $x$ . Now (9.12) is the monotone likelihood condition in terms of Savage [3]. Thus it follows from Theorem 6.1 of Savage [3] that if the condition (9.12) or (9.11) above is satisfied,

$$(9.13) \quad P(\mathbf{a}') \geq P(\mathbf{a})$$

for  $\mathbf{a}$  and  $\mathbf{a}'$  in (9.1) and (9.2) assuming  $i < j$ .

Now since  $\mathbf{b}$  in (8.9) is uniquely determined by  $\mathbf{a}$ , we may consider the statistic  $\phi(\mathbf{a}, \mathbf{b})$  in (8.14) as a function of  $\mathbf{a}$  alone, ignoring  $\mathbf{b}$ . Thus we can get a simple ordering of the vectors  $\mathbf{a}$ , according to the values of  $\phi(\mathbf{a}, \mathbf{b})$ . Similarly for a simple alternative  $G = \theta(F)$  we can have a simple ordering of vectors  $\mathbf{a}$ , according to the statistic  $P(\mathbf{a})$  in (3.7). Consider a fixed vector

$$\mathbf{a} = (\dots, a_i, \dots, a_j, \dots)$$

having  $a_i = a_j$ , ( $i < j$ ) and let  $\mathbf{a}'$  be the vector obtained from  $\mathbf{a}$  by replacing the  $i$ th and  $j$ th coordinates of  $\mathbf{a}$  by  $a_i + k$  and  $a_j - k$  respectively. For the same  $i$  and  $j$ , allowing  $k$  to take values  $1, 2, \dots, a_j$ , we get a set of vectors  $\mathbf{a}'$  which

we denote by  $\mathbf{S}'$ . Then we have from (9.1), (9.2), (9.4), (9.9), (9.11) and (9.13)

**THEOREM 9.** *For any function  $\theta$  satisfying (9.11), a simple ordering of  $\mathbf{S}'$  by the statistic  $P(\mathbf{a}')$  in (3.7) is identical with a simple ordering given by  $\phi(\mathbf{a}', \mathbf{b}')$  in (8.14).*

We shall have a similar theorem, if instead of (9.11) we have

$$(9.14) \quad \partial^2 G / \partial F^2 \leq 0.$$

Theorem 9 above also establishes a relation between the  $\phi$ -statistic in (8.14) and the  $V$ -statistic defined in (4.5) for which Theorem 7 was true.

**10. Numerical illustration.** In Table 3 we have given values of different statistics, for comparison. As already noted in Section 5, the ranking of the

TABLE 3  
 $m = 3 \quad n = 2$

1	2	3	4	5	6	7	8	9	10
$\mathbf{a}$	$G = F^{1/k}$ $k = 7.1663$ $P(\mathbf{a})$ (3.7)	$X$ -stat. (6.7)	$C_1$	$F, G$ Normal			$\phi$ -stat. (8.14)	$W$ -stat. (8.17)	Smirnov stat.
				equal var. $\delta =$ 0.75 Rel. freq. (Monte Carlo)	equal var. $\delta > 0$ $V$ -stat. (6.6)	equal means $\sigma_1 < \sigma_2$ $V$ -stat. (6.12)			
(3, 0, 0)	.0038	-1.40	-1.66	2.25	-1.68	1.41	1.00	1.00	1.00
(2, 1, 0)	.0054	-0.97	-1.16	3.45	-1.09	0.77	0.44	0.22	0.66
(1, 2, 0)	.0093	-0.54	-0.66	4.45	-0.59	0.77	0.44	0.22	0.50
(2, 0, 1)	.0073	-0.43	-0.50	5.45	-0.50	0.12	0.44	0.44	0.66
(1, 1, 1)	.0128	-0.00	-0.00	7.40	-0.00	0.12	0.22	0.11	0.33
(0, 3, 0)	.0667	0.00	0.00	8.10	0.00	1.41	0.50	0.50	0.50
(1, 0, 2)	.0214	0.43	0.50	11.15	0.50	0.12	0.44	0.44	0.66
(0, 2, 1)	.0919	0.54	0.66	12.00	0.59	0.77	0.44	0.22	0.50
(0, 1, 2)	.1537	0.97	1.16	18.45	1.09	0.77	0.44	0.22	0.66
(0, 0, 3)	.6277	1.40	1.66	27.30	1.68	1.41	1.00	1.00	1.00

vectors  $\mathbf{a}$  according to the probabilities  $P(\mathbf{a})$  in column 2 remains the same for all values of  $k > 1$  in Savage's Table [3]. Column 5 gives the relative frequencies in Teichroew's experiments [4]. Again the ranking of  $\mathbf{a}$ 's according to the relative frequencies, for all values of  $\delta$  considered in [4], remains nearly the same.

The ranking of vectors  $\mathbf{a}$  in column 1, by the  $V$ -statistic (for testing the variance ratio) in column 7 agrees better with the ranking by the  $\phi$  statistic in column 8 than with the rankings by the statistics in columns 9 and 10 respectively. It should be noted that the statistics  $X$ ,  $C_1$ , and  $V$  in columns 3, 4 and 6 respectively are meant to test one-sided alternatives. We can however construct, *intuitively*, two sided tests based on them, having the corresponding critical regions  $|X| > \text{const.}$ ,  $|C_1| > \text{const.}$  and  $|V| > \text{const.}$  Now in Table 3, the ranking of the vectors  $\mathbf{a}$  in column 1 by any of the statistics  $|X|$ ,  $|C_1|$ ,

and  $|V|$  agrees better with the ranking by the  $\phi$  statistic in column 8 than with the rankings by the statistics in columns 9 or 10 respectively. Of course more empirical investigation is necessary to arrive at practically usable conclusions.

**11. Some possibilities for the asymptotic behavior of the  $V$  and  $\phi$ -statistics.** This section consists of a few *conjectures* or *guesses*. From (4.5) we have,

$$(11.1) \quad \log V = \sum_1^n \log \theta' \left( \frac{1 + a_1 + \cdots + a_i}{m + 2} \right).$$

Now fixing the second sample  $y_1, \dots, y_n$ , let the size  $m$  of the first sample go to  $\infty$ . Then in view of (4.3), for both null and alternative hypotheses, with probability 1,

$$(11.2) \quad \log V = \sum_1^n \log \theta'(F(y_i)).$$

Now the asymptotic normality of  $\log V$  in (11.1) as  $n \rightarrow \infty$  could possibly be derived from the fact that the  $F(y)$ 's in (11.2) are distributed identically and independently on both the null and alternative hypotheses. On the null hypothesis the  $F(y)$ 's are distributed rectangularly,  $0 \leq F \leq 1$ . On the alternative hypothesis, the frequency function of  $F(y)$  is  $\theta'(F)$ ,  $0 \leq F \leq 1$ . This also suggests that the mean values of the asymptotic distribution of  $\log V$  in (11.1) might be

$$(11.3) \quad n \int_0^1 (\log \theta'(F)) dF,$$

$$(11.4) \quad n \int_0^1 (\log \theta'(F)) \theta'(F) dF$$

on the null and alternative hypotheses respectively. Similarly the variances, on the null and alternative hypotheses, could possibly be expressed as follows.

$$(11.5) \quad n \left\{ \int_0^1 (\log \theta'(F))^2 dF - \left[ \int_0^1 \log \theta'(F) dF \right]^2 \right\}$$

$$(11.6) \quad n \left\{ \int_0^1 (\log \theta'(F))^2 \theta'(F) dF - \left[ \int_0^1 (\log \theta'(F)) \theta'(F) dF \right]^2 \right\}.$$

The above integrals could be evaluated by means of numerical integration.

Next it seems from (4.3) that when  $m \rightarrow \infty$ ,  $n$  being fixed, the power of the  $V$ -test is equal to that of the corresponding optimum parametric test. In particular van der Waerden's  $X$ -test (6.7), which is equivalent to the corresponding  $V$ -test (6.6) as  $m \rightarrow \infty$ , has been proved in [5] to be asymptotically as powerful as the  $t$ -test.

The asymptotic distribution of the  $\phi$ -statistic in (8.14) is difficult to guess. However  $\phi(a)$  in (8.3) seems to be relatively easy to handle. From (8.3) we have



$$(11.7) \quad \log \phi(\mathbf{a}) = \log m! - m \log m + \sum_1^{n+1} a_i \log a_i - \sum_1^{n+1} \log a_i !.$$

Now suppose  $\mathbf{a} = (a_1, \dots, a_i, \dots, a_{n+1})$  in (11.7) is such that all the  $a_i$ 's are large enough so that Stirling's approximation can be applied to  $a_i !$ ,  $i = 1, \dots, n+1$ . Then from (11.7) we have

$$(11.8) \quad \log \phi(\mathbf{a}) = \text{const.} - \frac{1}{2} \sum_1^{n+1} \log a_i.$$

Further (except for degenerate alternatives) all the  $a_i$  will be large enough for Stirling's approximation, with as large a probability as we may wish, if, fixing the second sample  $y_1, \dots, y_n$ , we increase the size  $m$  of the first sample sufficiently. The asymptotic normality of the expressions in (11.7) and (11.8), ignoring constants, follows from a theorem due to Wolfowitz [6], under the condition  $m = n+1$ . Otherwise the asymptotic normality of (11.8) is obtainable from arguments similar to those in the preceding paragraph.

**Acknowledgment.** The author acknowledges with pleasure several helpful comments on the present work, by G. A. Barnard, Allan Birnbaum, E. L. Lehmann, J. L. Hodges, A. R. Kamat, S. S. Shrikhande, S. K. Mitra, C. M. Deo, the referee, and a reader. The author also thanks Mr. Modak for doing the necessary typing.

#### REFERENCES

- [1] WASSILY Hoeffding, "Optimum nonparametric tests," pp. 88-92 of *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951.
- [2] ERICH L. LEHMANN, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 22-43.
- [3] I. RICHARD SAVAGE, "Contribution to the theory of rank order statistics: Two sample case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 590-615.
- [4] D. TEICHROEW, "Empirical power functions for non-parametric two sample tests for small samples," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 340-344.
- [5] H. R. VAN DER WAERDEN, "Order tests for the two sample problem," *Koninkl. Nederl. Akademie van Wetenschappen*, Ser. A, Vol. 56, No. 4 (1954), pp. 303-316.
- [6] J. WOLFOWITZ, "Additive partition functions and a class of statistical hypotheses," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 247-279.

# A NONPARAMETRIC TEST FOR THE PROBLEM OF SEVERAL SAMPLES

BY V. P. BHAPKAR

*University of North Carolina and University of Poona*

**1. Summary.** In this paper, a new nonparametric test for the problem of  $c$  samples is offered. It is based upon the numbers of  $c$ -plets that can be formed by choosing one observation from each sample such that the observation from the  $i$ th sample is the least,  $i = 1, 2, \dots, c$ . The asymptotic distribution of the new test statistic is derived by an application of the extension of Hoeffding's theorem [4] on  $U$ -statistics to the case of  $c$  samples. The asymptotic power and the asymptotic efficiencies of this test relative to the Kruskal-Wallis  $H$ -test [7] and the Mood-Brown  $M$ -test [10] are computed in standard fashion along the lines of Andrews' paper [1].

**2. Introduction.** Let  $x_{i1}, x_{i2}, \dots, x_{in_i}$  be independent (real-valued) observations from the  $i$ th population with c.d.f.  $F_i$ ,  $i = 1, 2, \dots, c$ , and suppose that these  $c$  samples are independent. The  $F$ 's are assumed to be continuous. We consider a certain nonparametric test for the hypothesis

$$K_0: F_1 = F_2 = \dots = F_c.$$

If we assume that the populations are approximately of the same form, in the sense that if they differ it is by a shift or translation, then we may say that we are testing for the equality of location parameters. References to prior work on several-sample tests and some of the recent work may be found in [2], [6], [7], [8], and [10].

Let  $v^{(i)}$  be the number of  $c$ -plets that can be formed by choosing one observation from each sample such that the observation from the  $i$ th sample is the least. Then

$$(2.1) \quad v^{(i)} = \sum_{j=1}^{n_i} \prod_{r \neq i} \{\text{number of } x_{rs} > x_{ij}, \quad s = 1, 2, \dots, n_r\}.$$

The new test-statistic proposed is

$$(2.2) \quad V = N(2c - 1) \left[ \sum_{i=1}^c p_i(u^{(i)} - c^{-1})^2 - \left\{ \sum_{i=1}^c p_i(u^{(i)} - c^{-1}) \right\}^2 \right],$$

where  $N = \sum_i n_i$ ,  $p_i = n_i/N$  and  $u^{(i)} = v^{(i)} / (n_1 n_2 \dots n_c)$ . When the hypothesis  $K_0$  is true, it will be seen that the expectation of each  $u^{(i)}$  is  $1/c$ . Thus,  $V$  may be considered as a measure of deviation from  $K_0$ . The motivation behind the use of the  $v$ 's is simply to generalize, to the case of several samples, the Wilcoxon

Received September 10, 1959; revised April 27, 1961.

[12] statistic for two samples (the number of times observations in the first sample are smaller than observations in the second sample). The test consists in rejecting  $K_0$  at a significance level  $\alpha$  if  $V$  exceeds some predetermined number  $V_\alpha$ . In the next section it is shown that, when  $K_0$  is true,  $V$  is asymptotically distributed as a  $\chi^2$  variable with  $c - 1$  degrees of freedom. Thus, a large sample approximation for  $V_\alpha$  is provided by the upper  $\alpha$ -point of the  $\chi^2$  distribution with  $c - 1$  degrees of freedom. It is conjectured that this approximation is relatively close even for samples of moderate size.

**3. The asymptotic distribution of  $V$  under  $K_0$ .** It will be seen that

$$(3.1) \quad v = \sum_{i=1}^{(i)} \sum_{t_1=1}^{n_1} \sum_{t_2=1}^{n_2} \cdots \sum_{t_c=1}^{n_c} \phi^{(i)}(x_{1t_1}, x_{2t_2}, \dots, x_{ct_c}),$$

where

$$(3.2) \quad \phi^{(i)}(x_{1t_1}, x_{2t_2}, \dots, x_{ct_c}) = \begin{cases} 1 & \text{if } x_{it_i} < x_{kt_k} \text{ for all } k = 1, \dots, c \text{ except } i \\ 0 & \text{otherwise.} \end{cases}$$

Thus,  $u^{(i)}$  is a generalized  $U$ -statistic [11] corresponding to  $\phi^{(i)}$ . We shall make use of the following generalization of Hoeffding's theorem [4] on  $U$ -statistics to the case of  $c$  samples:

**LEMMA 3.1.** Let  $X_{ij}$ ,  $j = 1, 2, \dots, n_i$  for a fixed  $i$  be independent (real or vector) random variables identically distributed with c.d.f.  $F_i$ ,  $i = 1, 2, \dots, c$ . Further, let  $\sum_i n_i = N$  and

$$U_N^{(r)} = \left[ \prod_{i=1}^c \binom{n_i}{m_i^{(r)}} \right]^{-1} \sum^* \phi^{(r)}(X_{1\alpha_1}, \dots, X_{1\alpha_{m_1^{(r)}}}; \\ X_{2\beta_1}, \dots, X_{2\beta_{m_2^{(r)}}}; \dots; X_{c\delta_1}, \dots, X_{c\delta_{m_c^{(r)}}}), \quad r = 1, 2, \dots, g,$$

where each  $\phi^{(r)}$  is a function symmetric in each set of its arguments and  $\sum^*$  denotes the sum over all combinations  $(\alpha_1, \dots, \alpha_{m_1^{(r)}})$  of  $m_1^{(r)}$  integers chosen from  $(1, 2, \dots, n_1)$  and so on for  $\beta$ 's,  $\dots$ , and  $\delta$ 's. Assume that  $E[\phi^{(r)}] = \eta^{(r)}$  and  $E[\phi^{(r)}]^2 < \infty$ . Then

$$(i) \quad E[U_N^{(r)}] = \eta^{(r)},$$

$$(ii) \quad \text{Cov}[U_N^{(r)}, U_N^{(s)}] = \left[ \prod_{i=1}^c \binom{n_i}{m_i^{(s)}} \right]^{-1} \sum_{d_1=0}^{m_1^{(rs)}} \cdots \sum_{d_c=0}^{m_c^{(rs)}} \\ \cdot \prod_{i=1}^c \binom{m_i^{(r)}}{d_i} \binom{n_i - m_i^{(r)}}{m_i^{(s)} - d_i} \zeta_{d_1, d_2, \dots, d_c}(r, s),$$

where  $m_i^{(rs)} = \min(m_i^{(r)}, m_i^{(s)})$  and

$$(3.3) \quad \zeta_{d_1, d_2, \dots, d_c}(r, s) = \mathcal{E}[\phi^{(r)}(X_{11}, \dots, X_{1d_1}, X_{1d_1+1}, X_{1m_1^{(r)}}; \dots; X_{c1}, \dots, X_{cd_c}, X_{cd_c+1}, \dots, X_{cm_c^{(r)}}) \times \phi^{(s)}(X_{11}, \dots, X_{1d_1}, X_{1m_1^{(r)}+1}, \dots, X_{1m_1^{(r)}+m_1^{(s)}-d_1}; \dots; X_{c1}, \dots, X_{cd_c}, X_{cm_c^{(r)}+1}, \dots, X_{cm_c^{(r)}+m_c^{(s)}-d_c})] - \eta^{(r)} \eta^{(s)},$$

$r, s = 1, 2, \dots, g$ , it being understood that  $r = s$  gives us  $\text{Var}(U_N^{(r)})$ , and

(iii)  $N^{1/2}[\mathbf{U}_N - \mathbf{n}]$  is, in the limit as  $N \rightarrow \infty$  in such a way that  $n_i = Np_i$ , the  $p$ 's being fixed numbers such that  $\sum_i p_i = 1$ , normally distributed with zero mean and asymptotic covariance matrix  $\Sigma = (\sigma_{rs})$  given by

$$(3.4) \quad \sigma_{rs} = \sum_{i=1}^c \frac{m_i^{(r)} m_i^{(s)}}{p_i} \zeta_{0, \dots, 0, 1, 0, \dots, 0}^{(i)}(r, s), \quad r, s = 1, 2, \dots, g,$$

(1 at the  $i$ th place),

where  $\mathbf{U}'_N = (U_N^{(1)}, \dots, U_N^{(g)})$  and  $\mathbf{n}' = (\eta^{(1)}, \dots, \eta^{(g)})$ .

PROOF. The proof of this lemma (concerning generalized  $\mathbf{U}$ -statistics [11]) is a straightforward extension of the proof of Hoeffding's theorem [4] on  $U$ -statistics, and the details are omitted.

Now to apply the lemma to our problem, we note from (3.1) that

$$u^{(i)} = v^{(i)} / n_1 n_2 \cdots n_c, \quad i = 1, 2, \dots, c,$$

are generalized  $U$ -statistics with  $g = c$  and  $m_1^{(i)} = m_2^{(i)} = \dots = m_c^{(i)} = 1$ . Then if  $K_0$  is true,  $\eta^{(i)} = P[X_i < X_k \text{ for } k = 1, \dots, c \text{ except } k = i]$ , where the  $X$ 's are independent and identically distributed random variables, and hence

$$(3.5) \quad \begin{aligned} \eta^{(i)} &= c^{-1}; \\ \zeta_{0, \dots, 0, 1, 0, \dots, 0}^{(i)}(i, i) &= \mathcal{E}[\phi^{(i)}(X_1, \dots, X_i, \dots, X_c) \phi^{(i)}(X'_1, \dots, X_i, \dots, X'_c)] - c^{-2}, \end{aligned}$$

(1 at the  $i$ th place)

where again the  $X$ 's and  $X'$ 's are independent and identically distributed random variables, so that

$$(3.6) \quad \begin{aligned} \zeta_{0, \dots, 0, 1, 0, \dots, 0}^{(i)}(i, i) &= P[X_i < X_k, X_i < X'_k \text{ all } k = 1, \dots, c \text{ except } k = i] - c^{-2} \\ &= \frac{(c-1)^2}{c^2(2c-1)}; \end{aligned}$$

(1 at the  $i$ th place)

$$(3.7) \quad \begin{aligned} \zeta_{0, \dots, 0, 1, 0, \dots, 0}^{(i)}(i, i) &= P[X_i < X_k, X'_i < X_j, X'_i < X_l \text{ for all } k = 1, \dots, c \\ &\quad \text{except } i \text{ and all } l = 1, \dots, c \text{ except } i \text{ and } j] - c^{-2} \\ &= [c^2(2c-1)]^{-1}; \end{aligned}$$

(1 at the  $j$ th place)

and similarly,

$$(3.8) \quad \xi_{0, \dots, 0, 1, 0, \dots, 0}^{(i, j)} = \begin{cases} -\frac{c-1}{c^2(2c-1)} & \text{if 1 is at the } i\text{th or the } j\text{th place in the row of 0's} \\ [c^2(2c-1)]^{-1} & \text{otherwise.} \end{cases}$$

Thus, if  $K_0$  is true, from (3.4) we have

$$(3.9) \quad \sigma_{ii} = [c^2(2c-1)]^{-1} \left[ \frac{(c-1)^2}{p_i} + \sum_{k \neq i} \frac{1}{p_k} \right],$$

and

$$(3.10) \quad \sigma_{ij} = [c^2(2c-1)]^{-1} \left[ \left( \sum_k \frac{1}{p_k} \right) - \frac{c}{p_i} - \frac{c}{p_j} \right], \quad i \neq j.$$

The above two relations give us

$$(3.11) \quad c^2(2c-1)\Sigma = (\sum_k 1/p_k)J_{c,c} + c^2D - c\mathbf{q}J_{1,c} - cJ_{c,1}\mathbf{q}',$$

where  $D$  = diagonal  $(1/p_k, k = 1, 2, \dots, c)$ ,  $\mathbf{q}' = (1/p_1, \dots, 1/p_c)$  and  $J_{r,s} = (1)_{r,s}$ . Hence from (iii) in Lemma 3.1 it follows that  $N^{1/2}[\mathbf{U} - J_{c,1}/c]$ , where  $\mathbf{U}' = (u^{(1)}, \dots, u^{(c)})$ , has a limiting normal distribution with zero means and asymptotic covariance matrix  $\Sigma$  given by (3.11). But  $\sum_i v^{(i)} = n_1 n_2 \dots n_c$ , and hence  $u$ 's are subject to one linear constraint, viz.,  $\sum_i u^{(i)} = 1$ . Thus the distribution of  $u$ 's is singular and hence the asymptotic distribution is also singular. Then  $\Sigma$  is singular; in fact it can be easily verified from (3.11) that  $J_{1,c}\Sigma = \mathbf{0}$ . Let

$$N^{1/2}[\mathbf{U}' - J_{1,c}/c] = \mathbf{b}' = (b_1, \dots, b_{c-1}, b_c) = (\mathbf{b}'_0, b_c).$$

Then it follows that  $\mathbf{b}'_0 \Sigma_0^{-1} \mathbf{b}_0$  has a limiting  $\chi^2$  distribution with  $c-1$  degrees of freedom, where  $\Sigma_0$  denotes the asymptotic covariance matrix of  $\mathbf{b}_0$ . From (3.11) we have

$$(3.12) \quad c^2(2c-1)\Sigma_0 = aJ_{c-1,c-1} + c^2D_0 - c\mathbf{q}_0J_{1,c-1} - cJ_{c-1,1}\mathbf{q}'_0,$$

where  $D_0$  = diagonal  $(1/p_k, k = 1, 2, \dots, c-1)$ ,  $\mathbf{q}'_0 = (1/p_1, \dots, 1/p_{c-1})$  and  $a = \sum_{k=1}^c 1/p_k$ .

CASE (i):  $n_1 = n_2 = \dots = n_c$ . Then  $p_i = 1/c$  and (3.12) gives  $(2c-1)\Sigma_0 = c\mathbf{I} - J_{c-1,c-1}$ , so that

$$\Sigma_0^{-1} = (2c-1/c)[\mathbf{I} + J_{c-1,c-1}],$$

and hence,

$$(3.13) \quad \mathbf{b}'_0 \Sigma_0^{-1} \mathbf{b}_0 = \frac{N(2c-1)}{c} \sum_{i=1}^c \left( u^{(i)} - \frac{1}{c} \right)^2.$$

CASE (ii): Not all  $n$ 's are equal. Then  $q_0$  and  $J_{c-1,1}$  are linearly independent and from (3.12) we have

$$c^2(2c-1)\Sigma_0 = c^2D_0 - EF',$$

where

$$\Sigma = [cq_0, J_{c-1,1}], \quad F = [J_{c-1,1}, cq_0 - aJ_{c-1,1}]$$

are both of full rank viz., two. Then

$$[c^2(2c-1)]^{-1}\Sigma_0^{-1} = c^{-2}D_0^{-1} - D_0^{-1}E\Lambda F'D_0^{-1},$$

where  $\Lambda$  is given by

$$c^2[F'D_0^{-1}E - c^2I]\Lambda = I.$$

After simplification we finally have

$$(3.14) \quad b_0' \Sigma_0^{-1} b_0 = N(2c-1) \left[ \sum_{i=1}^c p_i \left( u^{(i)} - \frac{1}{c} \right)^2 - \left\{ \sum_{i=1}^c p_i \left( u^{(i)} - \frac{1}{c} \right) \right\}^2 \right].$$

It may be seen that the above expression reduces to (3.13) when  $n_1 = n_2 = \dots = n_c$ . It may be noted that the above expression is invariant under any choice of  $(c-1)$  linearly independent  $u$ 's. We have thus proved Theorem 3.1.

**THEOREM 3.1.** *If  $F_1 = F_2 = \dots = F_c$  and  $n_i = N_{p_i}$ , where the  $p$ 's are fixed numbers such that  $\sum_i p_i = 1$ , then the statistic  $V$ , defined by (2.2), has a limiting  $\chi^2$  distribution with  $c-1$  degrees of freedom as  $N \rightarrow \infty$ .*

**4. Consistency of the  $V$ -test.** As mentioned earlier, if we assume that the populations are approximately of the same form, then we may say that we are testing for the equality of location parameters. Thus, we are primarily interested in translation-type alternatives  $F_i(x) = F(x - \theta_i)$ ,  $i = 1, 2, \dots, c$ , where the  $\theta$ 's are not all equal. We shall show that the  $V$ -test is consistent against this class of alternatives.

We first state, without proof, the following straightforward extensions of a lemma of Lehmann ([9], p. 169).

**LEMMA 4.1.** *Let  $\eta = f(F_1, F_2, \dots, F_c)$  be a real-valued function such that  $f(F, F, \dots, F) = \eta_0$  for all  $(F, F, \dots, F)$  in a class  $\mathcal{C}_0$ . Let*

$$T_{n_1, \dots, n_c} = t(X_{11}, \dots, X_{1n_1}; \dots; X_{c1}, \dots, X_{cn_c})$$

*be a sequence of real-valued statistics such that  $T_{n_1, \dots, n_c}$  tends to  $\eta$  in probability as  $\min(n_1, \dots, n_c) \rightarrow \infty$ . Suppose that  $f(F_1, F_2, \dots, F_c) \neq \eta_0 (> \eta_0)$  for all  $(F_1, F_2, \dots, F_c)$  in a class  $\mathcal{C}_1$ . Then the sequence of tests which reject when  $|T_{n_1, \dots, n_c} - \eta_0| > c_{n_1, \dots, n_c}$  (when  $T_{n_1, \dots, n_c} - \eta_0 > c'_{n_1, \dots, n_c}$ ) is consistent for testing  $H: \mathcal{C}_0$  at every fixed level of significance against the alternatives  $\mathcal{C}_1$ .*

**LEMMA 4.2.** *Let  $\eta^{(i)} = f^{(i)}(F_1, F_2, \dots, F_c)$ ,  $i = 1, 2, \dots, g$ , be real-valued functions such that  $f^{(i)}(F, F, \dots, F) = \eta_0^{(i)}$  for all  $(F, F, \dots, F)$  in a class  $\mathcal{C}_0$ . Let  $T_{n_i, \dots, n_c}^{(i)} = t^{(i)}(X_{11}, \dots, X_{1n_1}; \dots; X_{c1}, \dots, X_{cn_c})$ ,  $i = 1, 2, \dots, g$ , be*



sequences of real-valued statistics such that  $T_{n_1}^{(i)}, \dots, n_c$  tends to  $\eta^{(i)}$  in probability as  $\min(n_1, \dots, n_c) \rightarrow \infty$ . Suppose that at least one  $f^{(i)}(F_1, F_2, \dots, F_c) \neq \eta_0^{(i)}$  for all  $(F_1, F_2, \dots, F_c)$  in a class  $\mathcal{C}_1$ . Further, let

$$W_{n_1, \dots, n_c} = \omega(T_{n_1}^{(1)}, \dots, T_{n_1}^{(g)}; T_{n_1}^{(g)}, \dots, T_{n_1}^{(g)})$$

be a nonnegative function which is zero if, and only if,  $T_{n_1}^{(i)} = \eta_0^{(i)}$  for all  $i = 1, \dots, g$ . Then the sequence of tests which reject when

$$W_{n_1, \dots, n_c} > d_{n_1, \dots, n_c}$$

is consistent for testing  $H: \mathcal{C}_0$  at every fixed level of significance against the alternatives  $\mathcal{C}_1$ .

If we take  $\eta^{(i)} = P[X_i < X_j \text{ for all } j = 1, \dots, c \text{ except } i]$ , where the  $X$ 's are independent random variables with continuous c.d.f.  $F_1, F_2, \dots, F_c$ , respectively, and  $T_{n_1}^{(i)}, \dots, n_c = u^{(i)}, i = 1, \dots, c$ , then the convergence in probability of  $u^{(i)}$  to  $\eta^{(i)}$  follows from (iii) in Lemma 3.1. For the class  $\mathcal{C}_1$  of translation-type alternatives  $F_i(x) = F(x - \theta_i)$ , where the  $\theta$ 's are not all equal, it may be easily seen that  $\eta^{(i)} > 1/c$ , where  $\theta_r$  is the (or one of the) least among  $\theta_1, \dots, \theta_c$ . The  $V$ -test, thus, is seen to be consistent against the class of translation-type alternatives.

More generally, the  $V$ -test is consistent against the wider class of alternatives for which  $P[X_i < X_j \text{ for all } j = 1, \dots, c \text{ except } i] \neq 1/c$  for at least one  $i$  among  $(1, \dots, c)$ , where the  $X$ 's are independent random variables with continuous c.d.f.  $F_1, F_2, \dots, F_c$ , respectively.

### 5. The asymptotic distribution of $V$ under translation-type alternatives.

Andrews [1] has investigated the asymptotic efficiencies of Kruskal's  $H$ -test and Mood's  $M$ -test and has concluded that the asymptotic efficiency of one relative to the other is  $\geq$  or  $\leq 1$ , for the translation-type alternatives, depending on the distribution function. It will be interesting (as suggested by Hoeffding and the referee) to carry out similar studies on this test with respect to the two previous tests. It is expected that the same type of conclusion will be reached.

Let us study the distribution of  $V$ , assuming a sequence of translation-type alternative hypotheses  $K_n$  for  $n = 1, 2, \dots$ . The hypothesis  $K_n$  specifies that  $F_i(x) = F(x - n^{-1}\theta_i)$ ,  $i = 1, 2, \dots, c$ , where not all  $\theta$ 's are equal. The letter  $n$  will be used to index a sequence of situations in which  $K_n$  is the true hypothesis. The limiting probability distribution will then be found as  $n \rightarrow \infty$ .

**THEOREM 5.1.** For each index  $n$  assume that  $n_i = ns_i$ , with  $s_i$  a positive integer and the truth of  $K_n$ .

If  $F$  possesses a continuous derivative  $f$  and there exists a function  $g$  such that

$$|f(y+h) - f(y)|/h \leq g(y)$$

and

$$\int_{-\infty}^{\infty} g(y)f(y) dy < \infty,$$

then, for  $n \rightarrow \infty$ , the statistic  $V$  has a limiting noncentral  $\chi^2$  distribution with  $c - 1$  degrees of freedom and the noncentrality parameter<sup>1</sup>

$$(5.1) \quad (2c - 1)c^2 \sum_{i=1}^c s_i(\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} [1 - F(y)]^{c-2} f^2(y) dy \right]^2,$$

where  $\bar{\theta} = \sum_i s_i \theta_i / \sum_i s_i$ .

PROOF: Let  $\eta_n^{(i)} = \mathcal{E}[\phi^{(i)}(X_1, X_2, \dots, X_c) | K_n]$ ; then it can be easily shown that

$$\eta_n^{(i)} = \frac{1}{c} - \frac{\delta_i}{n^{\frac{1}{2}}} \lambda + O(n^{-1}),$$

where

$$\delta_i = c\theta_i - \sum_{k=1}^c \theta_k,$$

and

$$(5.2) \quad \lambda = \int_{-\infty}^{\infty} [1 - F(y)]^{c-2} f^2(y) dy.$$

Similarly, it may be shown that

$$(5.3) \quad \Sigma_n = \Sigma + O(n^{-1}),$$

where  $\Sigma$  is given by (3.11) and  $O(n^{-1})$  denotes a matrix whose elements are  $O(n^{-1})$ .

Then, in view of Lemma 3.1,  $N^{\frac{1}{2}}(\mathbf{U} - \mathbf{n}_n)$  is, in the limit as  $n \rightarrow \infty$ , distributed with zero means and covariance matrix  $\Sigma_n$ , or, in view of (5.3), with asymptotic covariance matrix  $\Sigma$ . Hence  $N^{\frac{1}{2}}(\mathbf{U} - c^{-1}\mathbf{J}_{c,1})$  has a limiting normal distribution with mean-vector  $-(\sum_k s_k)^{\frac{1}{2}}\lambda\mathbf{\delta}$ , where  $\mathbf{\delta}' = (\delta_1, \dots, \delta_c)$ , and covariance matrix  $\Sigma$ . Thus  $V$ , in the limit as  $n \rightarrow \infty$ , is distributed as a noncentral  $\chi^2$  with  $c - 1$  degrees of freedom and the noncentrality parameter

$$\lambda_V = (\sum_k s_k)\lambda^2 \mathbf{\delta}_0' \Sigma_0^{-1} \mathbf{\delta}_0,$$

in the notation of Section 3. Since  $\sum_k \delta_k = 0$ , arguing exactly as from (3.11) to (3.14), we see that  $\lambda_V$  reduces to (5.1).

**6. Asymptotic relative efficiency.** Andrews [1] has shown that the  $H$ -statistic, the  $M$ -statistic and the  $F$ -statistic are asymptotically distributed as noncentral  $\chi^2$  with  $c - 1$  degrees of freedom and noncentrality parameters  $\lambda_H$ ,  $\lambda_M$  and  $\lambda_F$ , respectively, where

$$\lambda_H = 12 \left\{ \int_{-\infty}^{\infty} F'(x) dF(x) \right\}^2 \sum_{i=1}^c s_i(\theta_i - \bar{\theta})^2,$$

$$\lambda_M = 4[F'(a)]^2 \sum_{i=1}^c s_i(\theta_i - \bar{\theta})^2,$$

<sup>1</sup> This was also obtained independently by Y. S. Sathe.

and

$$\lambda_F = \sum_i s_i [(\theta_i - \bar{\theta})/\sigma_F]^2,$$

where  $a$  is the median of  $F$ .

It is now well known ([1], [3]) that in such cases the asymptotic efficiency of one statistic relative to the other is equal to the ratio of their noncentrality parameters. Hence, we have the asymptotic efficiencies of the  $V$ -statistic relative to the  $H$ ,  $M$  and  $F$  statistics as follows:

$$\epsilon_{V,H} = (2c - 1) c^2 \lambda^2 / 12 \left\{ \int_{-\infty}^{\infty} F'(x) dF(x) \right\}^2,$$

$$\epsilon_{V,M} = (2c - 1) c^2 \lambda^2 / 4 [F'(a)]^2,$$

and

$$\epsilon_{V,F} = (2c - 1) c^2 \lambda^2 \sigma_F^2,$$

respectively, where  $\lambda$  is given by (5.2). These expressions are seen to be independent of the scale parameter. For the uniform distribution the efficiencies are given by

$$\epsilon_{V,H} = \epsilon_{V,F} = \epsilon_{V,M}/3 = (2c - 1) c^2 / 12 (c - 1)^2,$$

so that we have

$c$	2	3	4	5	6	10	$\infty$
$\epsilon_{V,H}$	1.00	0.94	1.04	1.17	1.32	1.95	$\infty$

For the exponential distribution,  $f(y) = e^{-y}$ ,  $0 \leq y < \infty$ ,  $\epsilon_{V,H} = \epsilon_{V,M}/3 = \epsilon_{V,F}/3 = (2c - 1)/3$ , so that

$c$	2	3	4	5	10	$\infty$
$\epsilon_{V,H}$	1.00	1.66	2.33	3.00	6.33	$\infty$

For the normal distribution  $\lambda$  can be computed from the Table I given by Hojo [5] for  $c \leq 13$ . We have

$c$	2	3	4	5	6	7	8	10	12	13
$\epsilon_{V,H}$	1.00	0.94	0.86	0.80	0.74	0.69	0.65	0.58	0.53	0.51

while  $\epsilon_{V,M} = 3\epsilon_{V,H}/2$  and  $\epsilon_{V,F} = 3\epsilon_{V,H}/\pi$ .

For the normal distribution, the asymptotic efficiency of the  $V$ -statistic relative to the Kruskal-Wallis  $H$ -statistic tends to zero as the number of populations tends to infinity. I am thankful to the referee for supplying the following indication of the proof.

OUTLINE OF THE PROOF. We must show that

$$n^{\frac{1}{2}} \int_{-\infty}^{\infty} [\Phi(x)]^n \varphi^2(x) dx \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

On integrating by parts, it is seen that

$$\begin{aligned}\int_{-\infty}^{\infty} [\Phi(x)]^n \varphi^2(x) dx &= \frac{1}{n+1} \int_{-\infty}^{\infty} x [\Phi(x)]^{n+1} \varphi(x) dx \\ &= \frac{1}{n+1} \int_0^1 y^{n+1} \Phi^{-1}(y) dy.\end{aligned}$$

It is therefore enough to prove that

$$n^{\frac{1}{2}} \int_0^1 x^n \Phi^{-1}(x) dx \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We shall prove this using the fact that

$$\int_0^1 x^n [\log(x^{-1})]^{-1} dx = \left( \frac{\pi}{n+1} \right)^{\frac{1}{2}}.$$

It is easily seen by de l'Hospital's rule that

$$\frac{\Phi^{-1}(x)}{[\log(x^{-1})]^{-1}} \rightarrow 0 \quad \text{as } x \rightarrow 1.$$

Given any  $\epsilon > 0$  there exists therefore a constant  $a(\frac{1}{2} < a < 1)$  such that

$$\Phi^{-1}(x) \leq \epsilon [\log(x^{-1})]^{-1} \quad \text{for } a < x < 1,$$

and hence

$$\int_a^1 x^n \Phi^{-1}(x) dx \leq \epsilon \int_a^1 x^n [\log(x^{-1})]^{-1} dx \leq \epsilon \left( \frac{2\pi}{n+1} \right)^{\frac{1}{2}}.$$

Finally it is easily seen that for fixed  $a$   $\int_0^a x^n \Phi^{-1}(x) dx$  tends to 0 at a faster rate than  $\int_a^1 x^n \Phi^{-1}(x) dx$  as  $n \rightarrow \infty$ . Given  $\epsilon$  and hence  $a$ , there therefore exist  $n_0$  so that  $n \geq n_0$  implies

$$\int_0^1 x^n \Phi^{-1}(x) dx \leq 2 \int_a^1 x^n \Phi^{-1}(x) dx \leq 2\epsilon \left( \frac{2\pi}{n+1} \right)^{\frac{1}{2}}$$

and this completes the proof.

**7. Acknowledgment.** I wish to express my sincere thanks to Professor S. N. Roy and Professor Wassily Hoeffding for their keen interest in this work and their useful comments. I also wish to thank the referee for his suggestions.

#### REFERENCES

- [1] FRED C. ANDREWS, "Asymptotic behavior of some rank tests for analysis of variance," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 724-736.
- [2] MEYER DWASS, "Some  $k$ -sample rank-order tests," *Contributions to Probability and Statistics*, Stanford University Press, Stanford, California (1960), pp. 198-202.
- [3] E. J. HANNAN, "The asymptotic powers of certain tests based on multiple correlations," *J. Roy. Stat. Soc., Series B*, Vol. 18 (1956), pp. 227-233.
- [4] WASSILY Hoeffding, "A class of statistics with asymptotic normal distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 293-325.

- [5] TOKISHIGE HOJO, "Distribution of the median, quartiles and interquartile distance in samples from a normal population," *Biometrika*, Vol. 23 (1931), pp. 315-360.
- [6] J. KIEFER, "*K*-sample analogues of the Kolmogorov-Smirnov and Cramer-v. Mises tests," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 420-447.
- [7] WILLIAM H. KRUSKAL, "A nonparametric test for the several sample problem," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 525-540.
- [8] WILLIAM H. KRUSKAL AND W. ALLEN WALLIS, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 583-621.
- [9] E. L. LEHMANN, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165-179.
- [10] ALEXANDER MCFARLANE MOOD, *Introduction to the Theory of Statistics*, New York, McGraw-Hill Book Co., 1950.
- [11] BALKRISHNA V. SUKHATME, "Testing the hypothesis that two populations differ only in location," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 60-78.
- [12] FRANK WILCOXON, "Individual comparisons by ranking methods," *Biometrics Bull.*, Vol. 1 (1945), pp. 80-83.

# DISTRIBUTION OF THE ANDERSON-DARLING STATISTIC

BY PETER A. W. LEWIS

IBM Watson Laboratory, New York<sup>1</sup>

In [1] and [2] Anderson and Darling proposed the use of the statistic

$$(1) \quad W_n^2 = n \int_{-\infty}^{\infty} \frac{[G_n(x) - G(x)]^2}{G(x)[1 - G(x)]} dG(x)$$

for testing the hypothesis that a sample of size  $n$  has been drawn from a population with a specified continuous cumulative distribution function  $G(x)$ . In (1)  $G_n(x)$  is the empirical distribution function defined on the sample of size  $n$ .

We consider here the problem of determining and tabulating the distribution function,  $F(z; n) = \Pr \{W_n^2 \leq z\}$ , of this statistic. In [1], the asymptotic distribution of this statistic under the null hypothesis was derived and, rewritten in a form convenient for computation, it is given by

$$(2) \quad \begin{aligned} F(z; \infty) &= \lim_{n \rightarrow \infty} \Pr \{W_n^2 \leq z\} \\ &= \sum_{j=0}^{\infty} a_j (zb_j)^{\frac{1}{2}} \exp[-b_j/z] \int_0^{\infty} f_j(y) \exp[-y^2] dy, \end{aligned}$$

where

$$(3) \quad \begin{aligned} f_j(y) &= \exp[\frac{1}{2}zb_j/(y^2z + b_j)], \\ a_j &= \frac{(-1)^j (2)^{\frac{1}{2}} (4j+1) \Gamma(j + \frac{1}{2})}{j!}; \quad b_j = \frac{1}{8} (4j+1)^2 \pi^2. \end{aligned}$$

Using the calculated values of the  $a_j$ 's and  $b_j$ 's, and the fact that

$$\int_0^{\infty} f_j(y) e^{-y^2} dy \leq \frac{1}{2} (\pi)^{\frac{1}{2}} \exp[z/8],$$

it can be determined that no more than two terms of the sum ( $j = 0, 1$ ) are needed to evaluate  $F(z; \infty)$  to five decimal places over the range of  $z$  which is of interest. This range is  $0 \leq z \leq 8$ , since for all  $n$ ,  $F(8; n) = 1.000$ , rounded to three decimal places. The integral in each term of the sum was evaluated numerically using Hermite-Gauss quadrature numerical-integration formulas (p. 327 of [3], [4]). This method of numerical integration is very efficient in terms of computing time and gives sufficient accuracy to determine  $F(z; \infty)$  to five decimal places.

The results of these calculations of  $F(z; \infty)$ , rounded to four decimal places,

Received November 23, 1960; revised June 13, 1961.

<sup>1</sup> Present address: IBM Research Laboratory, Monterey and Cottle Roads, San Jose 14, California.



are given in the last column of Table I which appears at the end of this article. The asymptotic significance points given in [2] were verified and are shown in Table II.

An equivalent form of the statistic  $W_n^2$  is given by

$$(4) \quad W_n^2 =$$

$$-n - (n^{-1}) \sum_{i=1}^n [(2i-1)\ln G(X_{(i)}) + (2(n-i)+1)\ln(1-G(X_{(i)}))],$$

where the  $X_{(i)}$  are the order statistics of a sample of size  $n$ . It is well known that, under the null-hypothesis, the transformation  $G(X_{(i)})$  takes the  $X_{(i)}$  into the order statistics  $U_{(i)}$  of a sample of size  $n$  from a population with the uniform  $(0, 1)$  distribution, giving

$$(5) \quad W_n^2 = -n - (n^{-1}) \sum_{i=1}^n [(2i-1)\ln U_{(i)} + (2(n-i)+1)\ln(1-U_{(i)})].$$

This shows clearly the distribution-free property of this statistic under the null hypothesis, and allows us to determine very simply that, for any  $n$ , the minimum value which the random variable  $W_n^2$  can attain is

$$(6) \quad z(\min) = -n - \frac{1}{n} \sum_{i=1}^n \ln \left[ \left( \frac{2i-1}{2n} \right)^{2i-1} \left( \frac{2(n-i)+1}{2n} \right)^{2(n-i)+1} \right].$$

These values are tabulated in Table II (following Table I), in the row entitled  $F(z) = 0$ .

From equation (5), we find that

$$(7) \quad W_1^2 = -1 - \ln[U(1-U)],$$

where  $U$  is uniform  $(0, 1)$ , so that

$$(8) \quad F(z, 1) = \Pr\{W_1^2 \leq z\} = \begin{cases} (1 - 4 \exp[-(z+1)])^{\frac{1}{2}}, & z > .38629 \\ 0, & z \leq .38629 \end{cases}$$

Values of  $F(z, 1)$ , rounded to three decimal places, are given in Table I.

For  $n \geq 2$  and finite, resort was had to synthetic sampling (Monte Carlo) methods on an IBM 704 Computer to determine the distribution function  $F(z; n) = \Pr\{W_n^2 \leq z\}$ . This is done, using equation (5), by artificially generating  $m$  samples of size  $n$  from a uniform distribution. The result of this process is an empirical distribution function,  $F_m(z; n)$ , which is used as an estimate of  $F(z; n)$ . The  $F_m(z; n)$  are tabulated in Table I for  $n = 2$  up to  $n = 8$ .

It is necessary to make a determination of the accuracy of these estimates  $F_m(z; n)$  of  $F(z; n)$  for a given  $m$ . This can be done in either of two ways, as follows:

(1). For very large  $m$ ,  $mF_m(z; n)$  is approximately normally distributed, with mean  $mF(z; n)$  and variance  $mF(z; n)(1 - F(z; n))$ . Therefore a confidence interval with confidence coefficient  $1 - \alpha$  for  $F(z; n)$ , at any point  $z$ ,

is given by

$$(9) \quad \{F_m(z; n) \pm \zeta_{\alpha/2}([F(z)(1 - F(z))]/m)^{1/2}\}.$$

In this expression  $\zeta_{\alpha/2}$  is the upper  $-\alpha/2$  point of the  $N(0, 1)$  distribution. In most tables the "error" estimate used is essentially the above confidence interval with confidence coefficient 0.6868, i.e.,  $\zeta_{\alpha/2} = 1$ . Now  $F(z; n)$  is unknown, but  $F(z; n)(1 - F(z; n))$  is maximum at  $F(z; n) = 0.5$ , so that to keep this "error" less than or equal to .0005 over the range of  $z$ , one requires an  $m = 10^6$ .

(2). Another means of evaluating the error is by using the Kolmogorov-Smirnov statistic, from which, for large  $m$ , we can say that we are 95 percent sure that  $F_m(z; n)$  will stay within  $1.36(m)^{-1/2}$  of the true distribution  $F(z; n)$  for all  $z$ , i.e., over the entire distribution. Therefore to make this statement for a deviation of .0005, we need  $m = 7.398 \times 10^6$ .

Unfortunately the time available for computation limited the value of  $m$  used in these computations to  $m = 10^5$  for  $n = 2$ , and to  $m = .25 \times 10^6$  for  $n = 3, 4, 5, 6, 7$ , and 8. Thus, using the Kolmogorov-Smirnov criterion, the values  $F_m(z; 2)$  given in Table I are within .00163 of  $F(z, 2)$  with probability 0.95, and for  $n = 3, 4, 5, 6, 7$ , and 8 the values  $F_m(z, n)$  are within .00326 of  $F(z; n)$  with probability 0.95.

Determination of the distribution of  $W_n^2$  for  $n > 8$  by Monte Carlo methods is prohibitive, since for  $n = 8$  and  $m = 250,000$ , six hours of computing time were required. This is quite indicative of the inefficiency and impracticability of simple Monte Carlo methods as a means of solving distribution theory problems when the entire distribution function is required with great accuracy. Furthermore, it is doubtful whether modified Monte Carlo methods ([5], [6], [7]) could be used to advantage here.

Fortunately the convergence of the distribution of  $W_n^2$  to its asymptotic distribution is quite rapid. Thus, from Table I the maximum deviation at the tabulated points between the asymptotic distribution and the distribution for  $n = 8$  is approximately 0.006. For  $F(z) \geq 0.8$ , which is of most interest, this difference is only 0.001, so that for practical purposes the asymptotic distribution can be used for  $n > 8$ .

Significance points for  $W_n^2$  are given in Table II, for significance levels 0.100, 0.050, 0.010. For  $n = 1$  and  $n \rightarrow \infty$  these values are exact; the others are obtained by inverse interpolation from Table I and are only approximate.

**Acknowledgments.** I wish to thank Mr. H. Serenson for his help in programming this problem.

#### REFERENCES

- [1] T. W. ANDERSON AND D. A. DARLING, "Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193-212.
- [2] T. W. ANDERSON AND D. A. DARLING, "A test of goodness of fit," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 765-769.

- [3] F. B. HILDEBRAND, *Introduction to Numerical Analysis*, McGraw-Hill, New York, 1956.
- [4] H. E. SALZER, R. ZUCKER AND R. CAPUANO, "Tables of the zeros and weight factors of the first twenty Hermite polynomials," *J. of Res. of Natl. Bur. Stan.*, Vol. 48 (1952), pp. 111-116.
- [5] J. M. HAMMERSLEY AND K. W. MORTON, "A new Monte Carlo technique: antithetic variates," *Proc. Camb. Philos. Soc.*, Vol. 52 (1956), pp. 449-475.
- [6] J. M. HAMMERSLEY AND J. G. MAULDON, "General principles of antithetic variates," *Proc. Camb. Philos. Soc.*, Vol. 52 (1956), pp. 476-481.
- [7] J. M. HAMMERSLEY, "Conditional Monte Carlo," *J. Assn. Comp. Mach.*, Vol. 3 (1956), pp. 73-76.

*Tables follow on pages 1122-4*

TABLE I - Values of  $F(z; n)$  - Exact for  $n = 1$  and  $n \rightarrow \infty$ ; Estimated for  $n = 2, 3, 4, 5, 6, 7$ , and  $8$ 

$z \backslash n$	1	2	3	4	$F(z; n)$				8	$n \rightarrow \infty$
					5	6	7			
.025										0.0000
.050										0.0000
.075										0.0000
.100										0.0000
.125						0.000	0.000		0.000	0.0003
.150					0.000	0.001	0.001		0.001	0.0014
.175				0.001	0.003	0.003	0.003		0.004	0.0042
.200			0.008	0.007	0.008	0.009	0.008		0.009	0.0096
.225			0.016	0.016	0.016	0.017	0.017		0.017	0.0180
.250		0.001	0.028	0.028	0.028	0.029	0.029		0.029	0.0296
.275		0.030	0.044	0.043	0.044	0.044	0.044		0.045	0.0443
.300		0.059	0.063	0.063	0.063	0.062	0.062		0.063	0.0618
.325		0.087	0.083	0.085	0.083	0.084	0.083		0.084	0.0817
.350		0.115	0.106	0.109	0.106	0.106	0.106		0.106	0.1036
.375		0.142	0.130	0.134	0.130	0.131	0.130		0.130	0.1269
.400	0.116	0.169	0.159	0.161	0.156	0.156	0.155		0.155	0.1513
.425	0.195	0.196	0.187	0.187	0.182	0.182	0.181		0.181	0.1764
.450	0.248	0.222	0.217	0.212	0.208	0.208	0.207		0.207	0.2019
.475	0.291	0.248	0.248	0.238	0.235	0.234	0.233		0.233	0.2276
.500	0.328	0.273	0.271	0.264	0.261	0.260	0.259		0.259	0.2532
.525	0.360	0.298	0.295	0.289	0.287	0.285	0.284		0.284	0.2786
.550	0.389	0.323	0.320	0.314	0.312	0.310	0.309		0.309	0.3036
.575	0.415	0.347	0.345	0.340	0.337	0.335	0.334		0.334	0.3281
.600	0.439	0.371	0.371	0.364	0.361	0.359	0.358		0.358	0.3520
.625	0.461	0.394	0.396	0.387	0.384	0.382	0.381		0.381	0.3753
.650	0.481	0.418	0.418	0.410	0.407	0.404	0.403		0.404	0.3980
.675	0.501	0.440	0.439	0.431	0.429	0.426	0.424		0.425	0.4199
.700	0.519	0.463	0.459	0.452	0.449	0.446	0.446		0.446	0.4412
.750	0.552	0.507	0.496	0.491	0.489	0.486	0.486		0.487	0.4815
.800	0.582	0.547	0.530	0.528	0.525	0.524	0.523		0.523	0.5190
.850	0.609	0.580	0.567	0.563	0.559	0.559	0.557		0.557	0.5537
.900	0.634	0.610	0.598	0.593	0.591	0.590	0.588		0.589	0.5858
.950	0.656	0.636	0.626	0.622	0.620	0.619	0.618		0.619	0.6154
1.000	0.677	0.660	0.652	0.648	0.647	0.646	0.645		0.646	0.6427
1.050	0.696	0.683	0.676	0.673	0.672	0.671	0.669		0.670	0.6680
1.100	0.714	0.703	0.698	0.696	0.694	0.695	0.693		0.694	0.6912
1.150	0.731	0.722	0.719	0.717	0.715	0.716	0.714		0.714	0.7127
1.200	0.746	0.739	0.738	0.736	0.734	0.735	0.734		0.733	0.7324
1.250	0.761	0.756	0.755	0.754	0.752	0.753	0.752		0.751	0.7508
1.300	0.774	0.770	0.770	0.770	0.768	0.770	0.768		0.769	0.7677
1.350	0.786	0.784	0.785	0.785	0.784	0.785	0.784		0.784	0.7833
1.400	0.798	0.798	0.799	0.799	0.798	0.799	0.798		0.798	0.7978
1.450	0.809	0.809	0.811	0.812	0.811	0.812	0.812		0.811	0.8111
1.500	0.820	0.821	0.823	0.824	0.824	0.824	0.824		0.824	0.8235
1.550	0.829	0.831	0.833	0.835	0.835	0.835	0.835		0.835	0.8350
1.600	0.838	0.842	0.843	0.845	0.845	0.845	0.846		0.846	0.8457
1.650	0.847	0.851	0.852	0.855	0.854	0.855	0.855		0.855	0.8556
1.700	0.855	0.860	0.861	0.864	0.864	0.864	0.864		0.864	0.8648
1.750	0.863	0.868	0.869	0.872	0.872	0.872	0.873		0.873	0.8734
1.800	0.870	0.875	0.877	0.880	0.880	0.880	0.880		0.881	0.8814
1.850	0.877	0.883	0.884	0.887	0.887	0.887	0.888		0.888	0.8888
1.900	0.883	0.889	0.891	0.894	0.894	0.894	0.895		0.895	0.8957
1.950	0.889	0.896	0.898	0.900	0.901	0.900	0.901		0.901	0.9021
2.000	0.895	0.902	0.904	0.906	0.907	0.906	0.907		0.907	0.9082
2.050	0.900	0.907	0.909	0.912	0.912	0.912	0.913		0.913	0.9138
2.100	0.905	0.912	0.915	0.917	0.917	0.918	0.918		0.918	0.9190
2.150	0.910	0.917	0.920	0.922	0.922	0.923	0.922		0.923	0.9239

TABLE I (Continued)

$z$	$F(z; n)$								
	$n$	1	2	3	4	5	6	7	8
2.200		0.915	0.922	0.924	0.926	0.927	0.927	0.927	0.927
2.250		0.919	0.926	0.928	0.931	0.931	0.931	0.931	0.931
2.300		0.923	0.930	0.933	0.935	0.935	0.935	0.935	0.935
2.350		0.927	0.934	0.937	0.939	0.939	0.939	0.939	0.939
2.400		0.931	0.938	0.940	0.942	0.942	0.942	0.942	0.942
2.450		0.934	0.941	0.943	0.945	0.945	0.946	0.946	0.945
2.500		0.938	0.944	0.947	0.948	0.949	0.949	0.949	0.949
2.550		0.941	0.948	0.950	0.951	0.951	0.952	0.952	0.952
2.600		0.944	0.950	0.953	0.954	0.954	0.955	0.955	0.954
2.650		0.947	0.953	0.955	0.957	0.957	0.957	0.957	0.957
2.700		0.949	0.956	0.958	0.959	0.959	0.959	0.960	0.959
2.750		0.952	0.958	0.960	0.961	0.961	0.961	0.962	0.962
2.800		0.954	0.960	0.962	0.964	0.964	0.964	0.964	0.964
2.850		0.957	0.962	0.964	0.965	0.965	0.965	0.966	0.966
2.900		0.959	0.964	0.966	0.967	0.967	0.967	0.968	0.968
2.950		0.961	0.966	0.968	0.969	0.969	0.969	0.970	0.969
3.000		0.963	0.968	0.970	0.971	0.971	0.971	0.971	0.971
3.050		0.965	0.970	0.972	0.972	0.972	0.972	0.973	0.973
3.100		0.966	0.971	0.973	0.974	0.974	0.974	0.975	0.974
3.150		0.968	0.973	0.975	0.975	0.975	0.975	0.976	0.976
3.200		0.970	0.974	0.976	0.977	0.977	0.977	0.977	0.977
3.250		0.971	0.975	0.978	0.978	0.978	0.978	0.978	0.978
3.300		0.973	0.977	0.979	0.979	0.979	0.979	0.979	0.979
3.350		0.974	0.978	0.980	0.980	0.980	0.981	0.981	0.981
3.400		0.975	0.979	0.981	0.981	0.981	0.982	0.982	0.982
3.450		0.976	0.980	0.982	0.982	0.982	0.983	0.983	0.983
3.500		0.978	0.981	0.983	0.983	0.983	0.983	0.984	0.984
3.550		0.979	0.982	0.984	0.984	0.984	0.984	0.984	0.985
3.600		0.980	0.983	0.985	0.985	0.985	0.985	0.985	0.986
3.650		0.981	0.984	0.986	0.986	0.986	0.986	0.986	0.987
3.700		0.982	0.985	0.986	0.986	0.987	0.986	0.987	0.987
3.750		0.983	0.986	0.987	0.987	0.987	0.987	0.987	0.988
3.800		0.983	0.986	0.988	0.988	0.988	0.988	0.988	0.989
3.850		0.984	0.987	0.988	0.988	0.989	0.988	0.989	0.989
3.900		0.985	0.988	0.989	0.989	0.989	0.989	0.989	0.990
3.950		0.986	0.988	0.990	0.989	0.990	0.990	0.990	0.990
4.000		0.986	0.989	0.990	0.990	0.990	0.990	0.990	0.991
4.050		0.987	0.990	0.990	0.990	0.991	0.991	0.991	0.991
4.100		0.988	0.990	0.991	0.991	0.991	0.991	0.991	0.992
4.150		0.988	0.991	0.991	0.991	0.992	0.992	0.992	0.992
4.200		0.989	0.991	0.992	0.992	0.992	0.992	0.992	0.993
4.250		0.989	0.992	0.992	0.992	0.993	0.992	0.993	0.993
4.300		0.990	0.992	0.993	0.993	0.993	0.993	0.993	0.993
4.350		0.991	0.992	0.993	0.993	0.993	0.993	0.993	0.994
4.400		0.991	0.992	0.993	0.993	0.994	0.993	0.993	0.994
4.500		0.992	0.994	0.994	0.994	0.994	0.994	0.994	0.995
4.600		0.993	0.994	0.995	0.995	0.995	0.995	0.995	0.995
4.700		0.993	0.995	0.995	0.995	0.995	0.995	0.995	0.996
4.800		0.994	0.995	0.996	0.996	0.996	0.996	0.996	0.996
4.900		0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996
5.000		0.995	0.996	0.996	0.997	0.996	0.996	0.997	0.997
5.500		0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998
6.000		0.998	0.999	0.998	0.999	0.999	0.999	0.999	0.999
7.000		0.998	1.000	0.999	0.999	0.999	1.000	0.999	0.999
8.000		0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.999

TABLE II - Significance Points and Values of  $s(\min.)$ 

$$s = F^{-1} (F(u))$$

$F(u)$ \ $n$	1	2	3	4	5	6	7	8	$n \rightarrow \infty$
0	0.3863	0.2493	0.1885	0.1533	0.1304	0.1135	0.1043	0.0911	0
.90	2.0470	1.98	1.97	1.95	1.94	1.95	1.94	1.94	1.933
.95	2.7142	2.60	2.55	2.53	2.53	2.52	2.52	2.52	2.492
.99	4.3033	4.10	4.00	4.00	3.95	3.95	3.95	3.95	3.857



## ERRORS IN DISCRIMINATION

By S. JOHN

*Indian Statistical Institute, Calcutta*

**Summary.** The probabilities of misclassification involved in the use of estimated discriminant functions are subject to chance variations. The author's purpose in this paper is to derive the distribution laws that the probabilities of misclassification follow and to obtain their expected values. The parent populations are assumed to be normal. The first part of the paper considers the univariate case and the second part the multivariate case. The discussion of the multivariate case proceeds in three stages of increasing complexity. When the exact results are complicated, asymptotic results or approximations are given. Finally, the problem of estimating the expected probabilities of misclassification is considered. Interval estimates as well as point estimates are given.

**1. Introduction.** Multivariate statistical methods have been found extremely useful in devising efficient procedures for the solution of taxonomic problems. About twenty-five years ago Sir Ronald A. Fisher was consulted by M. M. Barnard as to the best method of classifying skeletal remains unearthed by archaeological excavations. Fisher suggested the use of the now well-known discriminant function [4], [7]. A general mathematical theory of statistical taxonomy was built by Welch [23] on foundations laid by Neyman and Pearson's theory of tests of hypotheses. Subsequent authors introduced many refinements. For a fairly complete account of the theory as it has developed during these years see chapter six of [3] or chapter eight of [18] and literature cited therein.

The situation we are considering is the following: We have an individual who has come from one of the two populations  $P^{(1)}$ ,  $P^{(2)}$ , but from which one is not known. It is required to devise a procedure that ensures a high probability of a correct classification of the individual. To come to a decision various characteristics of the individual are measured. Suppose we have measurements on  $p$  characteristics. Let the vector of measurements be  $\mathbf{x} = (x_1, x_2, \dots, x_p)$ . Let the distribution of these measurements in  $P^{(k)}$  have  $\boldsymbol{\mu}^{(k)}$  as its mean vector. Assume that the dispersion matrix is the same in both the populations. Denote this common dispersion matrix by  $\boldsymbol{\Sigma}$ . The discriminant function is then the linear function  $(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}\mathbf{x}'$ . We shall set

$$(1) \quad D(\mathbf{x}; \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}; \boldsymbol{\Sigma}) = (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}\mathbf{x}'.$$

The procedure usually adopted<sup>1</sup> is to classify the individual as belonging to  $P^{(1)}$  or  $P^{(2)}$  according as

$$(2) \quad D(\mathbf{x}; \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}; \boldsymbol{\Sigma}) \leq D(\tfrac{1}{2}[\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}]; \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}; \boldsymbol{\Sigma})$$

The above procedure is possible only if  $\boldsymbol{\mu}^{(k)}$  ( $k = 1, 2$ ) and  $\boldsymbol{\Sigma}$  are known. But

Received April 29, 1959; revised November 26, 1960.

<sup>1</sup> Certain situations require a slightly modified procedure. See Section 12.

usually such is not the case. We may then try to estimate the unknown parameters  $\mathbf{u}^{(1)}$ ,  $\mathbf{u}^{(2)}$  and  $\Sigma$  from random samples from  $P^{(1)}$  and  $P^{(2)}$ , substitute these estimates in the appropriate places and use the resulting function to classify individuals in exactly the same way as  $D(\mathbf{x}; \mathbf{u}^{(1)}; \mathbf{u}^{(2)}; \Sigma)$  is used.

Let  $x_{ir}^{(k)}$  ( $i = 1, 2, \dots, p; r = 1, 2, \dots, N_k$ ) be a random sample of size  $N_k$  from population  $P^{(k)}$ . Put

$$(3) \quad \bar{x}_i^{(k)} = N_k^{-1} \sum_{r=1}^{N_k} x_{ir}^{(k)} \quad i = 1, 2, \dots, p; k = 1, 2,$$

$$(4) \quad s_{ij} = (N_1 + N_2 - 2)^{-1} \sum_{k=1}^2 \sum_{r=1}^{N_k} [x_{ir}^{(k)} - \bar{x}_i^{(k)}][x_{jr}^{(k)} - \bar{x}_j^{(k)}], \quad i, j = 1, 2, \dots, p,$$

$$(5) \quad \bar{\mathbf{x}}^{(k)} = (\bar{x}_1^{(k)}, \bar{x}_2^{(k)}, \dots, \bar{x}_p^{(k)}) \quad k = 1, 2,$$

$$(6) \quad \mathbf{S} = (s_{ij}).$$

The vector  $\bar{\mathbf{x}}^{(k)}$  is an estimate of  $\mathbf{u}^{(k)}$  and the  $p \times p$  matrix  $\mathbf{S}$  is an estimate of the dispersion matrix  $\Sigma$ . Substituting these estimates in  $D(\mathbf{x}; \mathbf{u}^{(1)}, \mathbf{u}^{(2)}; \Sigma)$  we get  $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$ . Using this function we may assign individuals to  $P^{(1)}$  or  $P^{(2)}$  according as

$$(7) \quad (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}' \leq \frac{1}{2}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} + \bar{\mathbf{x}}^{(1)}).$$

In any classification procedure there are chances for two kinds of errors: (1) we may classify an individual from  $P^{(1)}$  as belonging to  $P^{(2)}$ ; (2) we may classify an individual from  $P^{(2)}$  as belonging to  $P^{(1)}$ . It is clear that if an individual is assigned to  $P^{(1)}$  or  $P^{(2)}$  depending on the value of a linear function  $\sum c_i x_i$ , these two chances will depend on the particular coefficients  $c_i$  used. Now, in  $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$ , the coefficients of  $x_1, x_2, \dots, x_p$  are respectively the components of the vector  $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}$ . These components are random variables. Random fluctuations in the coefficients induce random fluctuations in the chances of committing either kind of error and it is of interest to study these random fluctuations. This is what we do in the present paper. We assume  $P^{(k)}$  ( $k = 1, 2$ ) to be normal.

Wald's paper [22] appears to be the earliest one to discuss problems connected with the classification of an individual to  $P^{(1)}$  or  $P^{(2)}$ , when the distributions of the characteristics in  $P^{(1)}$  and  $P^{(2)}$  are not completely known. He considers the use of the statistic  $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$ . Wald had visualized a way of using this statistic slightly different from the one which we described. He required the distribution of  $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$  to set up the classification procedure. Papers [2], [8], [9], [10], [19] and [22] are partly or wholly concerned with the derivation of this distribution.<sup>2</sup> In [2], [10], [17] and [19], other statistics which can be used similarly are considered.

<sup>2</sup> A referee informs the author that Elfving has given an expansion for the unconditional probability in the univariate case and that Bowker and Sitgreaves have given an asymptotic expansion for the distribution function of the classification statistic when all parameters are estimated, in papers written for a forthcoming publication, *Mathematical Studies in Item Selection and Classification*, to be published by the Stanford University Press.

**2. Notation.** Besides the symbols already introduced in the introduction, we use other symbols also. We shall here explain the manner in which these symbols are to be construed.

To distinguish vectors and matrices from scalars we shall employ small bold face type to denote row vectors and capital bold face letters to denote matrices. The same letters, when primed, stand for the transposes of the vectors or matrices.

The letter **I** will denote the identity matrix of order  $p$ . If  $\mathbf{u} = (u_1, u_2, \dots, u_p)$ , we shall set

$$(9) \quad d\mathbf{u} = du_1 du_2 \cdots du_p.$$

The symbol  $g(x)$  will denote the standard normal density. The integral of  $g(x)$  from  $-\infty$  to  $x$  will be denoted by  $G(x)$ . The function inverse to  $G(x)$  will get the symbol  $G^{-1}(x)$ . We define

$$(9) \quad G(x_1, x_2; \rho) = \frac{(1 - \rho^2)^{-1/2}}{2\pi} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \exp[-\frac{1}{2}(1 - \rho^2)^{-1}(u_1^2 - 2\rho u_1 u_2 + u_2^2)] du_1 du_2.$$

The symbol  $I_x(p, q)$  will stand for the incomplete beta function,

$$(10) \quad \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x u^{p-1}(1-u)^{q-1} du.$$

Finally, we set

$$(11) \quad \delta^2 = (\mathbf{y}^{(2)} - \mathbf{y}^{(1)})\Sigma^{-1}(\mathbf{y}^{(2)} - \mathbf{y}^{(1)})'.$$

Besides the symbols introduced in this section, we use others locally. They will be explained at the appropriate places.

#### DISCRIMINATION USING A SINGLE CHARACTERISTIC

**3. Introduction to univariate case.** In the univariate case we shall for convenience write  $\mu^{(k)}$  for  $\mu_1^{(k)}$ ,  $\bar{x}^{(k)}$  for  $\bar{x}_1^{(k)}$ , and  $x$  for  $x_1$ .

It is easy to see that the general classification procedure described in the introduction reduces in the univariate case to the following: If  $\bar{x}^{(2)} > \bar{x}^{(1)}$ , assign the individual to  $P^{(1)}$  or  $P^{(2)}$  according as  $x \leq [\bar{x}^{(1)} + \bar{x}^{(2)}]/2$ . If  $\bar{x}^{(2)} \leq \bar{x}^{(1)}$ , assign the individual to  $P^{(1)}$  or  $P^{(2)}$  according as  $x \geq [\bar{x}^{(1)} + \bar{x}^{(2)}]/2$ .

Suppose  $\bar{x}^{(1)}$  and  $\bar{x}^{(2)}$  are given. We shall denote by  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$  the conditional probability of assigning an individual from  $P^{(1)}$  to  $P^{(2)}$  and by  $e_{21}(\bar{x}^{(1)}, \bar{x}^{(2)})$  the conditional probability of assigning an individual from  $P^{(2)}$  to  $P^{(1)}$ .

$$(12) \quad e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) = \begin{cases} 1 - G([\sigma^{(1)}]^{-1}[\frac{1}{2}\{\bar{x}^{(1)} + \bar{x}^{(2)}\} - \mu^{(1)}]) & \text{if } \bar{x}^{(1)} < \bar{x}^{(2)}, \\ G([\sigma^{(1)}]^{-1}[\frac{1}{2}\{\bar{x}^{(1)} + \bar{x}^{(2)}\} - \mu^{(1)}]) & \text{if } \bar{x}^{(1)} \geq \bar{x}^{(2)}. \end{cases}$$

Here  $\sigma^{(k)}$  denotes the standard deviation of the characteristic under consideration, in population  $P^{(k)}$ . A similar equation for  $e_{21}(\bar{x}^{(1)}, \bar{x}^{(2)})$  can be written down

at once. We shall obtain the distribution and expected value of  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ . A discussion of  $e_{21}(\bar{x}^{(1)}, \bar{x}^{(2)})$  would be completely analogous.

The classification procedure we have described is usually adopted only if  $\sigma^{(1)} = \sigma^{(2)}$ . However, in obtaining the distribution and expected value of  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$  we shall not assume that  $\sigma^{(1)} = \sigma^{(2)}$ ; there is some interest in studying the chances of errors under the more general set-up, since, although the classification procedure was designed on the assumption that  $\sigma^{(1)} = \sigma^{(2)}$ , there is a possibility that the assumption was false.

**4. The distribution of  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ .** The quantity  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$  can be less than  $z$  if and only if either of the following two events happen:

$$\bar{x}^{(1)} < \bar{x}^{(2)} \quad \text{and} \quad \frac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}] - \mu^{(1)} > -\sigma^{(1)}G^{-1}(z)$$

or

$$\bar{x}^{(1)} \geq \bar{x}^{(2)} \quad \text{and} \quad \frac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}] - \mu^{(1)} < \sigma^{(1)}G^{-1}(z).$$

The distribution function of  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$  is therefore given by the equation

$$(13) \quad \Pr(e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) < z) = G(h_{11}, h_{21}; \rho) + G(h_{12}, h_{22}; \rho),$$

where

$$(14) \quad h_{11} = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-1}(\mu^{(2)} - \mu^{(1)}) = -h_{12},$$

$$(15) \quad h_{21} = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-1}[2\sigma^{(1)}G^{-1}(z) + \mu^{(2)} - \mu^{(1)}],$$

$$(16) \quad h_{22} = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-1}[2\sigma^{(1)}G^{-1}(z) - \mu^{(2)} + \mu^{(1)}],$$

and

$$(17) \quad \rho = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-1}(N_2^{-1}[\sigma^{(2)}]^2 - N_1^{-1}[\sigma^{(1)}]^2).$$

The expression on the right hand side in equation (13) can be evaluated using the tables of  $G(x_1, x_2; \rho)$  given in [15]. If  $N_1^{-1}[\sigma^{(1)}]^2 = N_2^{-1}[\sigma^{(2)}]^2$ ,

$$(18) \quad \Pr(e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) < z) = G(h_{11})G(h_{21}) + G(h_{12})G(h_{22}),$$

and hence can be evaluated with the help of the tables of  $G(x)$  given in [14].

**5. Expected value of  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ .** The expected value of  $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$  can be calculated from the equation

$$(19) \quad Ee_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) = G(a_{11}, a_{21}; \rho) + G(a_{12}, a_{22}; \rho),$$

where

$$(20) \quad a_{11} = -(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-1}(\mu^{(2)} - \mu^{(1)}) = -a_{12},$$

$$a_{21} = \frac{1}{2}\{[\sigma^{(1)}]^2 + \frac{1}{4}(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)\}^{-1}(\mu^{(2)} - \mu^{(1)}) = -a_{22},$$

and

$$\rho = \frac{1}{2}(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-1}$$

$$(21) \quad \{[\sigma^{(1)}]^2 + \frac{1}{4}(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)\}^{-1} (N_1^{-1}[\sigma^{(1)}]^2 - N_2^{-1}[\sigma^{(2)}]^2).$$

If  $N_1^{-1}[\sigma^{(1)}]^2 = N_2^{-1}[\sigma^{(2)}]^2$ ,  $Ee_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$  can be evaluated using only tables of  $G(x)$ ; for, in this case,

$$(22) \quad Ee_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) = G(a_{11})G(a_{21}) + G(a_{12})G(a_{22}).$$

Equation (19) is easily established if we observe that a wrong assignment of an individual from  $P^{(1)}$  corresponds to the occurrence of either of the following two events:

$$\bar{x}^{(1)} < \bar{x}^{(2)} \quad \text{and} \quad x \geq \frac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}]$$

or,

$$\bar{x}^{(1)} \geq \bar{x}^{(2)} \quad \text{and} \quad x \leq \frac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}].$$

The reader may wish to compare our treatment of the univariate case with that of [11].

#### DISCRIMINATION USING MORE THAN ONE CHARACTERISTIC

**6. Introduction to the multivariate case.** We now take up for consideration the multivariate case. The procedure discussed is the one described in the introduction. It is an adaptation of the standard discriminant function analysis to situations where the parameters required for the construction of the discriminant function are unknown.

Classification procedures based on the correct discriminant function are known to be the best possible when the distributions in the two populations are multivariate normal with identical dispersion matrices. We shall, throughout our discussion, assume that the distributions in the two populations, do, in fact, satisfy these conditions.

The discussion will proceed in several stages. We shall, at stage number one, assume that only  $\mathbf{y}^{(2)}$  is unknown. The case where only  $\mathbf{y}^{(1)}$  is unknown is completely analogous and does not require separate consideration. At stage two we shall only assume that the dispersion matrix  $\Sigma$  is known. In the third stage we shall not assume that  $\mathbf{y}^{(1)}$ ,  $\mathbf{y}^{(2)}$ , or  $\Sigma$  are known.

**7. Case one: only  $\mu^{(2)}$  is unknown.** Before starting discussion of this case let us note that we shall not err seriously if we take  $\bar{\mathbf{x}}^{(1)}$  to be the true value of  $\mathbf{y}^{(1)}$  and  $\mathbf{S}$  to be the true value of  $\Sigma$ , provided  $N_1$  is sufficiently large.

For constructing the discriminant function,  $\mathbf{y}^{(2)}$  has to be estimated. Substituting  $\bar{\mathbf{x}}^{(2)}$  for  $\mathbf{y}^{(2)}$  we have the discriminant function,

$$(23) \quad D(\mathbf{x}; \mathbf{y}^{(1)}, \bar{\mathbf{x}}^{(2)}; \Sigma) = (\bar{\mathbf{x}}^{(2)} - \mathbf{y}^{(1)})\Sigma^{-1}\mathbf{x}.$$

An individual with measurements  $\mathbf{x}$  is assigned to  $P^{(1)}$  or  $P^{(2)}$  according as

$$(24) \quad D(\mathbf{x}; \mathbf{y}^{(1)}, \bar{\mathbf{x}}^{(2)}; \Sigma) \leq D(\frac{1}{2}[\mathbf{y}^{(1)} + \bar{\mathbf{x}}^{(2)}]; \mathbf{y}^{(1)}, \bar{\mathbf{x}}^{(2)}; \Sigma).$$

7.1. *Distribution of  $e_{12}(\bar{\mathbf{x}}^{(2)})$ .* Given  $\bar{\mathbf{x}}^{(2)}$ , the probability of misclassifying an individual from  $P^{(1)}$  is  $1 - G(y)$  where

$$(25) \quad y = \frac{1}{2}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})'.$$

We shall denote this probability by  $e_{12}(\bar{\mathbf{x}}^{(2)})$ . Clearly,  $e_{12}(\bar{\mathbf{x}}^{(2)})$  is a random variable since it depends on  $\bar{\mathbf{x}}^{(2)}$ . The distribution function of  $e_{12}(\bar{\mathbf{x}}^{(2)})$  is given by the equation

$$(26) \quad \Pr(e_{12}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr(4N_2 y^2 > 4N_2 [G^{-1}(z)]^2) \quad (0 \leq z \leq \frac{1}{2}).$$

Now,  $4N_2 y^2$  is a noncentral chisquare variable with  $p$  degrees of freedom and noncentrality<sup>3</sup> equal to  $(N_2 \delta^2)/2$ .  $\Pr(e_{12}(\bar{\mathbf{x}}^{(2)}) < z)$  can therefore be determined from tables of the noncentral chisquare distribution<sup>4</sup>.

It is interesting to note that  $\mathbf{u}^{(1)}$ ,  $\mathbf{u}^{(2)}$  and  $\Sigma$  enter into the distribution of  $e_{12}(\bar{\mathbf{x}}^{(2)})$  only in the form of  $\delta$ . For any given  $z$ ,  $\Pr(e_{12}(\bar{\mathbf{x}}^{(2)}) < z)$  is a monotonic function of  $\delta$  and therefore can be asserted to lie between certain bounds provided we know upper and lower bounds for  $\delta$ .

7.2. *Expected value of  $e_{12}(\bar{\mathbf{x}}^{(2)})$ .* From the preceding section we see that

$$(27) \quad e_{12}(\bar{\mathbf{x}}^{(2)}) = 1 - G(\frac{1}{2}[v/N_2]^{\frac{1}{2}})$$

where

$$(28) \quad v = N_2(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})'.$$

The random variable  $v$  has the density function

$$(29) \quad 2^{-\frac{1}{2}p} e^{-\frac{1}{2}(\lambda + v)} v^{\frac{1}{2}p-1} \sum_{r=0}^{\infty} [\Gamma(\frac{1}{2}p + r)]^{-1} \frac{(\frac{1}{2}\lambda v)^r}{r!},$$

where

$$(30) \quad \lambda = \frac{1}{2}N_2 \delta^2.$$

Therefore,

$$(31) \quad \begin{aligned} Ee_{12}(\bar{\mathbf{x}}^{(2)}) &= 2^{-\frac{1}{2}p} e^{-\lambda} \int_0^{\infty} v^{\frac{1}{2}p-1} e^{-\frac{1}{2}v} \left\{ \sum_{r=0}^{\infty} [\Gamma(\frac{1}{2}p + r)]^{-1} \frac{(\frac{1}{2}\lambda v)^r}{r!} \right\} \\ &\quad \left\{ \int_{\frac{1}{2}(v/N_2)^{\frac{1}{2}}}^{\infty} g(x) dx \right\} dv, \\ &= \frac{1}{2} e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} I_a(\frac{1}{2}p + r, \frac{1}{2}) \end{aligned}$$

where  $a = 4N_2/(1 + 4N_2)$ .

<sup>3</sup> Some authors use the term "noncentrality" for twice this number.

<sup>4</sup> Tables now available are not exactly in the form we require. Editors of [14] have announced that tables of the probability integral of the noncentral chisquare distribution are among the tables considered for inclusion in Vol. II. For the present, recourse must be had to approximate methods developed in [1] and [13].



The justification for the last step is the fact that

$$(32) \quad [\Gamma(\frac{1}{2}p + r)]^{-1} 2^{-\frac{1}{2}(p+2r)} \int_0^\infty v^{\frac{1}{2}p+r-1} e^{-\frac{1}{2}v} dv \int_{\frac{1}{2}(r/N_2)^{\frac{1}{2}}}^\infty g(x) dx$$

is equal to half the probability that a random variable having the  $F$ -distribution with degrees of freedom one and  $p + 2r$  takes a value greater than  $(4N_2)^{-1}(p + 2r)$ .

It is possible to give several other expressions for  $Ee_{12}(\bar{\mathbf{x}}^{(2)})$ ; the one we have given above appeared to be the most convenient.

7.3. *Distribution of  $e_{21}(\bar{\mathbf{x}}^{(2)})$ .* Thus far we have been discussing the chances of wrongly assigning an individual from  $P^{(1)}$  to  $P^{(2)}$ . We now take up consideration of the probability of wrongly assigning an individual from  $P^{(2)}$  to  $P^{(1)}$ .

Given  $\bar{\mathbf{x}}^{(2)}$ , the probability of misclassifying an individual from  $P^{(2)}$  is  $G(w)$  where

$$(33) \quad w = \frac{1}{2}[(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})']^{\frac{1}{2}} \\ - [(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})']^{-\frac{1}{2}}[(\mathbf{u}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})'].$$

This probability we denote by  $e_{21}(\bar{\mathbf{x}}^{(2)})$ . Obviously  $e_{21}(\bar{\mathbf{x}}^{(2)})$  is a random variable. We shall derive its distribution.

The distribution function of  $e_{21}(\bar{\mathbf{x}}^{(2)})$  is given by the equation

$$(34) \quad \Pr(e_{21}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr(w < G^{-1}(z)).$$

This equation shows that it suffices to derive the distribution of  $w$ .

Observe that  $w$  is a function of

$$(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})', \quad \text{and} \quad (\mathbf{u}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})'.$$

Set

$$(35) \quad t_1 = (\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})'$$

and

$$(36) \quad t_2 = (\mathbf{u}^{(2)} - \mathbf{u}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(1)})'.$$

Without loss of generality we may assume that  $\mathbf{u}^{(1)} = \mathbf{0}$  and  $\Sigma = \mathbf{I}$ . The density function of  $\bar{\mathbf{x}}^{(2)}$  is then

$$(37) \quad (N_2/2\pi)^{\frac{1}{2}p} \exp[-\frac{1}{2}N_2(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(2)})(\bar{\mathbf{x}}^{(2)} - \mathbf{u}^{(2)})'] \\ = (N_2/2\pi)^{\frac{1}{2}p} \exp[-\frac{1}{2}N_2(t_1 - 2t_2 + \delta^2)].$$

Therefore, if we denote by  $f(t_1, t_2)$  the joint density function of  $t_1$  and  $t_2$ ,

$$\begin{aligned}
 f(t_1, t_2) dt_1 dt_2 &= \int \cdots \int_{\substack{t_1 < \mathbf{x}^{(2)} \mathbf{x}^{(2)'} < t_1 + d t_1 \\ t_2 < \mathbf{y}^{(2)} \mathbf{y}^{(2)'} < t_2 + d t_2}} (N_2/2\pi)^{1/2} \exp[-\frac{1}{2}N_2(t_1 - 2t_2 + \delta^2)] d\mathbf{x}^{(2)}, \\
 (38) \quad &= (N_2/2\pi)^{1/2} \exp[-\frac{1}{2}N_2(t_1 - 2t_2 + \delta^2)] \int \cdots \int_{\substack{t_1 < \mathbf{x}^{(2)} \mathbf{x}^{(2)'} < t_1 + d t_1 \\ t_2 < \mathbf{y}^{(2)} \mathbf{y}^{(2)'} < t_2 + d t_2}} d\mathbf{x}^{(2)}, \\
 &= (N_2/2)^{1/2} \pi^{-1} [\delta \Gamma(\frac{1}{2}(p-1))]^{-1} \\
 &\quad \cdot [t_1 - (t_2/\delta)]^{1/2(p-3)} \exp[-\frac{1}{2}N_2(t_1 - 2t_2 + \delta^2)] dt_1 dt_2,
 \end{aligned}$$

since [21]

$$(39) \quad \int \cdots \int_{\substack{t_1 < \mathbf{x}^{(2)} \mathbf{x}^{(2)'} < t_1 + d t_1 \\ t_2 < \mathbf{y}^{(2)} \mathbf{y}^{(2)'} < t_2 + d t_2}} d\mathbf{x}^{(2)} = \pi^{1/2(p-1)} [\Gamma(\frac{1}{2}(p-1))]^{-1} \delta^{-1} [t_1 - (t_2/\delta)]^{1/2(p-3)} dt_1 dt_2.$$

Substituting

$$(40) \quad t_1 = u^2, \quad w = \frac{1}{2}t_1 - t_1^{-1}t_2,$$

we find that the joint density function of  $u$  and  $w$  is

$$(41) \quad C u^{p-1} [1 - \delta^{-2}(w - \frac{1}{2}u)^2]^{1/2(p-3)} \exp[-\frac{1}{2}N_2(2uw + \delta^2)]$$

where

$$(42) \quad C = 2^{-1/2(p-2)} \pi^{-1} [\delta \Gamma(\frac{1}{2}(p-1))]^{-1} N_2^{1/2}.$$

Integrating out  $u$  we obtain as the density function of  $w$  the function

$$(43) \quad h(w) = \begin{cases} C e^{-1/2(N_2 \delta^2)} \int_{\frac{1}{2}(w-\delta)}^{2(w+\delta)} u^{p-1} [1 - \delta^{-2}(w - \frac{1}{2}u)^2]^{1/2(p-3)} e^{-N_2 u w} du & \text{if } w \geq \delta, \\ C e^{-1/2(N_2 \delta^2)} \int_0^{2(w+\delta)} u^{p-1} [1 - \delta^{-2}(w - \frac{1}{2}u)^2]^{1/2(p-3)} e^{-N_2 u w} du & \text{if } -\delta \leq w \leq \delta, \\ 0 & \text{if } w < -\delta. \end{cases}$$

If  $p$  is odd, the expression within square brackets in the integrand may be expanded and each term integrated by parts. For example, if  $p = 3$ ,

$$(44) \quad h(w) = \begin{cases} (2N_2)^{1/2} \pi^{-1} (\delta w)^{-1} e^{-1/2(N_2 \delta^2)} [e^{-2N_2 w(w-\delta)} \{2(w-\delta)^2 \\ + 2(N_2 w)^{-1}(w-\delta) + (N_2 w)^{-2}\} \\ - e^{-2N_2 w(w+\delta)} \{2(w+\delta)^2 \\ + 2(N_2 w)^{-1}(w+\delta) + (N_2 w)^{-2}\}] & \text{if } w \geq \delta \\ (2N_2)^{1/2} \pi^{-1} (\delta w)^{-1} e^{-1/2(N_2 \delta^2)} [(N_2 w)^{-2} - e^{-2N_2 w(w+\delta)} \{2(w+\delta)^2 \\ + 2(N_2 w)^{-1}(w+\delta) + (N_2 w)^{-2}\}] & \text{if } -\delta \leq w < \delta \text{ but } w \neq 0 \\ 2^{1/2} 3^{-1} \pi^{-1} N_2^{1/2} e^{-1/2(N_2 \delta^2)} \delta^2 & \text{if } w = 0, \\ 0 & \text{if } w < -\delta. \end{cases}$$

For even values of  $p$  either recourse must be had to numerical integration or percentage points must be obtained by interpolation from corresponding percentage points for distributions with  $p$  an odd integer.

Here we observe that  $w > -\delta$  with probability one. Therefore, with probability one

$$(45) \quad e_{21}(\bar{\mathbf{x}}^{(2)}) > G(-\delta).$$

From equations (27) and (28) we see that  $e_{12}(\bar{\mathbf{x}}^{(2)})$ , on the other hand, can go down even to zero.

Observe that besides  $N_2$  and  $p$ ,  $\delta$  is the only parameter entering into  $h(w)$ .

7.4. *Asymptotic distribution of  $e_{21}(\bar{\mathbf{x}}^{(2)})$ .* Since the exact distribution of  $e_{21}(\bar{\mathbf{x}}^{(2)})$  is somewhat complicated, it may be useful to note that, as  $N_2 \rightarrow \infty$ , the distribution of  $2N_2^{1/2}[g(\delta/2)]^{-1}[e_{21}(\bar{\mathbf{x}}^{(2)}) - G(-\delta/2)]$  tends (weakly) to the normal distribution with mean zero and variance unity.

Perhaps it is better to use the asymptotic distribution of  $w$  together with equation (34). The limiting distribution of  $2N_2^{1/2}(w + \frac{1}{2}\delta)$  is normal with mean zero and unit variance. Hence we have, using equation (34),

$$(46) \quad \Pr(e_{21}(\bar{\mathbf{x}}^{(2)}) < z) \approx G(2N_2^{1/2}[G^{-1}(z) + \frac{1}{2}\delta]).$$

## 8. Case two: only $\Sigma$ is known.

8.1. *Introduction to Case two.* Since  $\mathbf{y}^{(1)}$  and  $\mathbf{y}^{(2)}$  are unknown we shall construct the discriminant function using  $\bar{\mathbf{x}}^{(1)}$  and  $\bar{\mathbf{x}}^{(2)}$ . The resulting discriminant function is

$$(47) \quad D(\mathbf{x}; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \Sigma) = (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\Sigma^{-1}\mathbf{x}'.$$

The classification procedure consists in assigning individuals to  $P^{(1)}$  or  $P^{(2)}$  according as

$$(48) \quad D(\mathbf{x}; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \Sigma) \leq D(\frac{1}{2}[\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}]; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \Sigma).$$

Given  $\bar{\mathbf{x}}^{(1)}$  and  $\bar{\mathbf{x}}^{(2)}$ , the probability of misclassifying an individual from  $P^{(1)}$  is  $1 - G(u_1)$  where

$$(49) \quad u_1 = \frac{1}{2}[(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})]^{1/2} + [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})]^{-1/2}[(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(1)} - \mathbf{y}^{(1)})].$$

We shall denote this by  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ . Similarly let  $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  denote the conditional probability of misclassifying an individual from  $P^{(2)}$ . Being functions of random variables,  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  and  $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  are themselves random variables. We shall obtain the distribution of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ . Since we are free to regard either of the two populations as  $P^{(1)}$  it is not necessary to consider  $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  separately.

8.2. *The distribution of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ .* Since

$$(50) \quad \Pr(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) < z) = \Pr(u_1 > -G^{-1}(z)),$$

the distribution function of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  will be determined when the distribution

of  $u_1$  is obtained. In deriving the distribution of  $u_1$  we shall assume that  $\Sigma = \mathbf{I}$  and  $\mathbf{u}^{(1)} = \mathbf{0}$ . Clearly, there is no loss of generality in doing so.

From equation (49) we observe that  $u_1$  is a function of  $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}$  and  $\bar{\mathbf{x}}^{(1)}$ . The joint distribution of  $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}$  and  $\bar{\mathbf{x}}^{(1)}$  is multivariate normal with means  $(\mathbf{u}^{(2)}, \mathbf{0})$  and variance-covariance matrix

$$(51) \quad \begin{pmatrix} [N_1^{-1} + N_2^{-1}]\mathbf{I} & -N_1^{-1}\mathbf{I} \\ -N_1^{-1}\mathbf{I} & N_1^{-1}\mathbf{I} \end{pmatrix}.$$

Therefore, the distribution of  $\bar{\mathbf{x}}^{(1)}$ , given  $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$ , is multivariate normal with mean vector

$$-(N_1 + N_2)^{-1}N_2(\mathbf{y} - \mathbf{u}^{(2)})$$

and dispersion matrix  $(N_1 + N_2)^{-1}\mathbf{I}$ . It follows, therefore, that, given  $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$ ,  $u_1$  has the normal distribution with mean

$$(52) \quad \begin{aligned} & \frac{1}{2}(N_1 + N_2)^{-1}(N_1 - N_2)(\mathbf{y}\mathbf{y}')^{\frac{1}{2}} + (N_1 + N_2)^{-1}N_2(\mathbf{u}^{(2)}\mathbf{y}')(\mathbf{y}\mathbf{y}')^{-\frac{1}{2}} \\ & = \frac{1}{2}(N_1 + N_2)^{-1}(N_1 - N_2)t_3^{\frac{1}{2}} + (N_1 + N_2)^{-1}N_2t_3^{-\frac{1}{2}}t_4 \quad (\text{say}) \end{aligned}$$

and variance  $(N_1 + N_2)^{-1}$ . Hence, if  $h(u_1)$  denotes the density function of  $u_1$  and  $f(t_3, t_4)$  the joint density function of  $t_3$  and  $t_4$ , we have the equation

$$(53) \quad \begin{aligned} h(u_1) = & \iint \left( \frac{N_1 + N_2}{2\pi} \right)^{\frac{1}{2}} \exp \left[ -\frac{N_1 + N_2}{2} \right. \\ & \left. \cdot \left( u_1 - \frac{1}{2} \frac{N_1 - N_2}{N_1 + N_2} t_3^{\frac{1}{2}} - \frac{N_2}{N_1 + N_2} t_3^{-\frac{1}{2}} t_4 \right)^2 \right] f(t_3, t_4) dt_3 dt_4 \end{aligned}$$

where the region of integration is the entire domain of variation of  $t_3$  and  $t_4$ . It is thus necessary to obtain the joint density function  $f(t_3, t_4)$  of  $t_3$  and  $t_4$ . This is done in the next section.

### 8.3. The joint density function of $t_3$ and $t_4$ .

$$(54) \quad t_3 = (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})' = \mathbf{y}\mathbf{y}';$$

$$(55) \quad t_4 = \mathbf{u}^{(2)}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})' = \mathbf{u}^{(2)}\mathbf{y}'.$$

The distribution of  $\mathbf{y}$  is multivariate normal with mean  $\mathbf{u}^{(2)}$  and variance-covariance matrix  $(N_1^{-1} + N_2^{-1})\mathbf{I}$ . Therefore,

$$\begin{aligned} f(t_3, t_4) dt_3 dt_4 = & (2\pi)^{-p/2} \left( \frac{N_1 N_2}{N_1 + N_2} \right)^{\frac{1}{2}p} \int \dots \int_{\substack{t_3 < \mathbf{y}\mathbf{y}' < t_3 + dt_3 \\ t_4 < \mathbf{u}^{(2)}\mathbf{y}' < t_4 + dt_4}} \\ & \cdot \exp \left[ -\frac{N_1 N_2}{2(N_1 + N_2)} (\mathbf{y} - \mathbf{u}^{(2)})(\mathbf{y} - \mathbf{u}^{(2)})' \right] d\mathbf{y} \end{aligned}$$

$$\begin{aligned}
 (56) \quad &= (2\pi)^{-p/2} \left( \frac{N_1 N_2}{N_1 + N_2} \right)^{1/2} \\
 &\cdot \exp \left[ -\frac{1}{2} \frac{N_1 N_2}{N_1 + N_2} (t_3 - 2t_4 + \delta^2) \right] \times \int \cdots \int_{\substack{t_3 < y < t_3 + \delta t_3 \\ t_4 < y^{(2)} < t_4 + \delta t_4}} dy \\
 &= (2\pi)^{-1/2} \left( \frac{N_1 N_2}{N_1 + N_2} \right)^{1/2} \\
 &\quad \cdot \exp \left[ -\frac{1}{2} \frac{N_1 N_2}{N_1 + N_2} (t_3 - 2t_4 + \delta^2) \right] \\
 &\quad \times \frac{\pi^{1/2(p-1)}}{\Gamma(1/2(p-1))} \cdot \delta^{-1} [t_3 - (t_4/\delta)^2]^{1/2(p-1)} dt_3 dt_4.
 \end{aligned}$$

8.4. *The distribution of  $u_1$  when  $N_1 = N_2$ .* If the two sample sizes  $N_1$  and  $N_2$  are equal to  $N$  (say), the distribution of  $u_1$  takes a simpler form. In this case, given  $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$ , the conditional distribution of  $u_1$  is normal with mean

$$(57) \quad \frac{1}{2} t_3^{-1} t_4 = \frac{1}{2} \delta t \quad (\text{say})$$

and variance  $(2N)^{-1}$ . Therefore, if  $f(t)$  is the density function of  $t$ , we may write

$$(58) \quad h(u_1) = \int_{-1}^1 (N/\pi)^{1/2} \exp[-N(u_1 - \frac{1}{2}\delta t)^2] f(t) dt.$$

It is now necessary to obtain the density function of  $t$ . For this we have only to use the joint density function  $f(t_3, t_4)$  of  $t_3$  and  $t_4$  to find the joint density of  $t_3$  and  $t$  and integrate out  $t_3$  from this joint density. The resulting expression for  $f(t)$  is given by the equation

$$(59) \quad f(t) = \pi^{-1} [\Gamma(\frac{1}{2}(p-1))]^{-1} e^{-1/2 N t^2} (1-t^2)^{1/2(p-3)} \sum_{r=0}^{\infty} \frac{\Gamma(\frac{1}{2}(p+r))}{r!} N^{1/2} (\delta t)^r.$$

8.5. *An alternative derivation of the distribution of  $u_1$  in the case  $N_1 = N_2$ .* In the general case we gave the distribution of  $u_1$  as a double integral. When  $N_1 = N_2$ , we are able to give the density function of  $u_1$  as a single integral. In this case it is also possible to give the density function of  $u_1$  in a more explicit form using a different method.

Let  $\varphi_{u_1}(\theta)$  be the characteristic function of  $u_1$ . We have seen that, given  $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$ ,  $u_1$  has the normal distribution with mean  $(\delta t/2)$  and variance  $(2N)^{-1}$ . Therefore, the conditional characteristic function is

$$(60) \quad \exp [i \delta t \theta - (4N)^{-1} \theta^2].$$

Hence we have

$$(61) \quad \varphi_{u_1}(\theta) = \int_{-1}^{+1} \exp [i \delta t \theta - (4N)^{-1} \theta^2] f(t) dt.$$

For the sake of convenience we now change over from  $u_1$  to the variable  $v_1$  defined by the equation

$$(62) \quad v_1 = (2N)^{1/2} u_1.$$

Let  $\varphi_{v_1}(\theta)$  denote the characteristic function of  $v_1$ . Then,

$$(63) \quad \begin{aligned} \varphi_{v_1}(\theta) &= \varphi_{u_1}([2N]^{1/2}\theta), \\ &= \int_{-1}^1 \exp[i(N/2)^{1/2}\delta t\theta - \theta^2/2] f(t) dt, \\ &= \pi^{-1} \{ \Gamma(\frac{1}{2}[p-1]) \}^{-1} \exp(-\frac{1}{4}N\delta^2 - \frac{1}{2}\theta^2) \\ &\quad \cdot \int_{-1}^1 \left[ \sum_{r=0}^{\infty} \frac{\Gamma(\frac{1}{2}[p+r])}{r!} N^{r/2} (\delta t)^r \right] (1-t^2)^{(p-3)/2} e^{i(N/2)^{1/2}\delta t\theta} dt. \end{aligned}$$

Expanding the exponential factor of the integrand and integrating term by term we obtain the following equation for  $\varphi_{v_1}(\theta)$ :

$$(64) \quad \varphi_{v_1}(\theta) = \pi^{-1} \exp(-\frac{1}{4}N\delta^2 - \frac{1}{2}\theta^2) \sum' \frac{\Gamma(\frac{1}{2}[p+r])\Gamma(\frac{1}{2}[r+m+1])}{\Gamma(\frac{1}{2}[p+r+m])} \cdot \frac{(N\delta^2)^{(r+m)/2} (-\frac{1}{2}\theta^2)^{m/2}}{r! m!}$$

where  $\sum'$  denotes summation over all non-negative integral values of  $r$  and  $m$  such that  $r+m$  is an even integer.

The inversion formula for characteristic functions now readily yields an expression for the density function of  $v_1$ . This expression is

$$(65) \quad \pi^{-1} e^{-N\delta^2/4 - v_1^2/4} \sum' \frac{\Gamma(\frac{1}{2}[p+r])\Gamma(\frac{1}{2}[r+m+1])}{\Gamma(\frac{1}{2}[p+r+m])} \cdot \frac{(N\delta^2)^{(r+m)/2} 2^{-(m-1)/2}}{r! m!} \cdot H_m(v_1).$$

Here  $H_r(x)$  denotes the Hermite polynomial of degree  $r$  defined by the equation

$$(66) \quad \left(-\frac{d}{dx}\right)^r e^{-x^2} = H_r(x) e^{-x^2}.$$

8.6. *The asymptotic distribution of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ .* As  $N_1$  and  $N_2$  tend to infinity, the distribution of  $2(N_1^{-1} + N_2^{-2})^{-1} [g(\frac{1}{2}\delta)]^{-1} [e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) - G(-\frac{1}{2}\delta)]$  tends weakly to the normal distribution with zero mean and unit variance.

Again it may be better to use the asymptotic distribution of  $u_1$  together with equation (50). The limiting distribution of  $2(N_1^{-1} + N_2^{-2})^{-1} (u_1 - \frac{1}{2}\delta)$  is normal with zero mean and unit variance. Hence we have

$$(67) \quad \Pr(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) < z) \approx G(2[N_1^{-1} + N_2^{-1}]^{-1} [G^{-1}(z) + \frac{1}{2}\delta]).$$

9. Case three:  $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}$  and  $\Sigma$  unknown. In this case the discriminant function



as constructed from sample estimates of  $\mathbf{y}^{(1)}$ ,  $\mathbf{y}^{(2)}$  and  $\Sigma$  is

$$(68) \quad D(\mathbf{x}; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) = (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'.$$

The classification procedure consists in assigning individuals to  $P^{(1)}$  or  $P^{(2)}$  according as

$$(69) \quad D(\mathbf{x}; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) \leq D(\frac{1}{2}[\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}]; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}).$$

Given  $\bar{\mathbf{x}}^{(1)}$ ,  $\bar{\mathbf{x}}^{(2)}$  and  $\mathbf{S}$ , the probability of misclassifying an individual from  $P^{(1)}$  is  $1 - G(w_1)$  where

$$(70) \quad w_1 = [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\Sigma\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{-\frac{1}{2}} \\ [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\frac{1}{2}[\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}] - \mathbf{y}^{(1)})'].$$

We shall denote this probability by  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S})$ . Clearly,  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S})$  is a random variable. The exact distribution of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S})$  is complicated and we shall be content with giving its asymptotic distribution. To be slightly more general, we shall suppose that  $\mathbf{S}$  is some estimate of  $\Sigma$  with  $n$  degrees of freedom and independent of  $\bar{\mathbf{x}}^{(1)}$  and  $\bar{\mathbf{x}}^{(2)}$ , not necessarily obtained exclusively from the same samples as those from which  $\bar{\mathbf{x}}^{(1)}$  and  $\bar{\mathbf{x}}^{(2)}$  were obtained. It can then be shown that as  $N_1$ ,  $N_2$  and  $n$  tend to infinity, the distribution of  $2[N_1^{-1} + N_2^{-1}]^{-\frac{1}{2}}[g(\frac{1}{2}\delta)]^{-1} [e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) + G(\frac{1}{2}\delta) - 1]$  tends (weakly) to the normal distribution with mean zero and unit variance. We have also, corresponding to equations (46) and (67) of previous sections, for large values of  $N_1$ ,  $N_2$  and  $n$ ,

$$(71) \quad \Pr(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) < z) \approx G(2[N_1^{-1} + N_2^{-1}]^{-\frac{1}{2}}[G^{-1}(z) + \frac{1}{2}\delta]).$$

## 10. Expected probability of misclassification.

10.1. *Introduction.* We have been discussing in previous sections sampling fluctuations in the chances of misclassification involved in using estimated discriminant functions. Distributions and expected values were the objects of investigation. The expected values were evaluated in certain special cases using *ad hoc* methods. We shall, in this section, give a unified treatment.<sup>6</sup> The method of this section is capable of yielding expected values in all cases where the dispersion matrix is known.

Besides exact expressions some simple approximations also will be given.

10.2. *Exact expressions.* Consider the case where only the variance-covariance matrix  $\Sigma$  is assumed to be known. We require expected values of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  and  $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ . We shall evaluate  $Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  only, since  $Ee_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  can be evaluated on similar lines.

Using results contained in [10], it is possible to prove that

$$(72) \quad Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = \Pr(v_2^{-1}v_1 > [1 + \rho]^{-1}[1 - \rho])$$

<sup>6</sup> Though the method of this section has the merit of greater generality, the expressions obtained in the earlier sections are somewhat simpler.

where  $v_1$  and  $v_2$  are independent non-central chisquare variables each having  $p$  degrees of freedom and non-centralities given respectively by

$$(73) \quad \lambda_1 = [4(1 + \rho)]^{-1} N_1 N_2 [(N_1 + N_2)^{-1} - (N_1 + N_2 + 4N_1 N_2)^{-1}]^2 \delta^2$$

and

$$(74) \quad \lambda_2 = [4(1 - \rho)]^{-1} N_1 N_2 [(N_1 + N_2)^{-1} + (N_1 + N_2 + 4N_1 N_2)^{-1}]^2 \delta^2.$$

Here

$$(75) \quad \rho = [(N_1 + N_2)(N_1 + N_2 + 4N_1 N_2)]^{-1} (N_2 - N_1).$$

Using the expression for the density function of  $v_2^{-1} v_1$  we may write

$$(76) \quad Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = e^{-(\lambda_1 + \lambda_2)} \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{\Gamma(p+r+s)}{\Gamma(\frac{1}{2}p+r)\Gamma(\frac{1}{2}p+s)} \cdot \frac{\lambda_1^r \lambda_2^s}{r! s!} \cdot \int_{(1-\rho)/(1+\rho)}^{\infty} (1+u)^{-(p+r+s)} u^{\frac{1}{2}p+r-1} du.$$

We shall put equation (76) in the slightly different form

$$(77) \quad Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = e^{-(\lambda_1 + \lambda_2)} \left[ \sum_{r=0}^{\infty} \sum_{s=0}^r \frac{\lambda_1^r \lambda_2^s}{r! s!} \{1 - I_{\frac{1}{2}(1-\rho)}(\frac{1}{2}p+r, \frac{1}{2}p+s)\} + \sum_{r=0}^{\infty} \sum_{s=r+1}^{\infty} \frac{\lambda_1^r \lambda_2^s}{r! s!} I_{\frac{1}{2}(1+\rho)}(\frac{1}{2}p+r, \frac{1}{2}p+s) \right].$$

The various terms on the right hand side in equation (77) can be evaluated using tables of  $I_x(p, q)$  given in [16].

In case  $\mathbf{y}^{(1)}$  also is known with  $\Sigma$ , we have only to take, in equation (77),

$$(78) \quad \begin{aligned} \lambda_1 &= [4(1 + \rho)]^{-1} N_2 [1 - (1 + 4N_2)^{-1}]^2 \delta^2, \\ \lambda_2 &= [4(1 - \rho)]^{-1} N_2 [1 + (1 + 4N_2)^{-1}]^2 \delta^2, \end{aligned}$$

and

$$(79) \quad \rho = -[1 + 4N_2]^{-1}$$

to get the expected value of  $e_{12}(\bar{\mathbf{x}}^{(2)})$ . If the known parameters are  $\Sigma$  and  $\mathbf{y}^{(2)}$ , we take

$$(80) \quad \begin{aligned} \lambda_1 &= [4(1 + \rho)]^{-1} N_1 [1 - (1 + 4N_1)^{-1}]^2 \delta^2, \\ \lambda_2 &= [4(1 - \rho)]^{-1} N_1 [1 + (1 + 4N_1)^{-1}]^2 \delta^2, \end{aligned}$$

and

$$(81) \quad \rho = (1 + 4N_1)^{-1}.$$

**10.3. An approximation.** In the previous section we obtained an exact expression for the expected value of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ . That expression is not quite convenient for numerical evaluation. For this reason we now give an approximation which permits evaluation of the expected value using only tables of  $G(x)$ .

We start with the result

$$(82) \quad Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = \Pr(v_2^{-1}v_1 > [1 + \rho]^{-1}[1 - \rho]).$$

Now,

$$(83) \quad \begin{aligned} \Pr(v_2^{-1}v_1 > [1 + \rho]^{-1}[1 - \rho]) &= \Pr(v_2^{-1}v_1^{\dagger} > [1 + \rho]^{-1}[1 - \rho]^{\dagger}), \\ &= \Pr([1 + \rho]^{\dagger}v_1^{\dagger} - [1 - \rho]^{\dagger}v_2^{\dagger} > 0). \end{aligned}$$

From [1] we know that if  $\chi^{2'}$  has a noncentral chisquare distribution with  $f$  degrees of freedom and noncentrality parameter equal to  $\lambda$ , the variable  $(\chi^{2'}/r)^{\dagger}$ , where

$$(84) \quad r = f + 2\lambda,$$

has approximately a normal distribution with expectation  $1 - 2(1 + b)/9r$  and variance  $2(1 + b)/9r$ . (Here  $b$  stands for  $2[f + 2\lambda]^{-1}\lambda$ ). Using this result and also equations (82) and (83) we may now write

$$(85) \quad Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) \approx G(a)$$

where

$$(86) \quad \begin{aligned} a &= (18)^{-\dagger} [r_1^{-\dagger}(1 + \rho)^{\dagger}(1 + b_1) + r_2^{-\dagger}(1 - \rho)^{\dagger}(1 + b_2)]^{-\dagger} \\ &\quad [2\{r_2^{-\dagger}(1 - \rho)^{\dagger}(1 + b_2) - r_1^{-\dagger}(1 + \rho)^{\dagger}(1 + b_1)\} \\ &\quad + 9\{r_1(1 + \rho)\}^{\dagger} - 9\{r_2(1 - \rho)\}^{\dagger}], \end{aligned}$$

the quantities  $r_1, r_2, b_1$  and  $b_2$  being defined by the equations

$$(87) \quad \begin{aligned} r_i &= p + 2\lambda_i & (i = 1, 2), \\ b_i &= 2(p + 2\lambda_i)^{-1}\lambda_i & (i = 1, 2). \end{aligned}$$

The question of the closeness of the approximation (85) now arises. We should expect that the approximation involved is of the same order as that involved in assuming that the cube root of a noncentral chisquare variable has the normal distribution. Though it would be interesting to compare the approximate values with the corresponding exact values using numerical computations, we shall not embark on this venture at the present moment. Some numerical results given in [1] may be found enlightening.

**11. Estimation of the expected probabilities of misclassification.** The expressions for the expected probabilities of misclassification are found to be functions of  $\delta$ . The problem of estimating these expected probabilities now arises. The empirical method of estimating the probability of an event by computing the proportion of outcomes favourable to the event in a number of repetitions of the experiment is available to us. This method is suggested in [20]. If the problem is one of estimating the conditional probabilities of misclassification, the empirical method is a simple way of solving it. What we have to do is to use the estimated discriminant function to classify the  $N_i$  individuals known to belong to  $P^{(i)}$

and note down the proportion of these individuals assigned to  $P^{(i)} (i = 1 \text{ and } j = 2 \text{ or } i = 2 \text{ and } j = 1)$ .

If the problem is one of estimating the unconditional probabilities of misclassification, the empirical method can prove exasperating. Fortunately, the maximum likelihood estimator is simple enough. To obtain the maximum likelihood estimate, all we have to do is to substitute  $(\bar{\mathbf{x}}^{(2)} - \mathbf{y}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{y}^{(1)})'$ , or  $(\mathbf{y}^{(2)} - \bar{\mathbf{x}}^{(1)})\Sigma^{-1}(\mathbf{y}^{(2)} - \bar{\mathbf{x}}^{(1)})'$ , or  $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$ , or

$$(N_1 + N_2 - 2)^{-1}(N_1 + N_2)(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$$

for  $\delta^2$  in the expressions for the expected error (or approximations to them), depending on which parameters are known.

The exact value of the expected probability is certain to differ from its estimate. An idea of the magnitude of the difference that may be expected can be obtained from the variance of the estimator. But the expression for the variance happens to be quite cumbersome. Besides, it involves unknown parameters. Fortunately another approach is open to us. We indicate below how intervals which enclose the true value of the probability with any preassigned degree of certainty can be constructed.

The procedure to be adopted is the following: Suppose the desired confidence level is  $\alpha$ . Set up for  $\delta$  a confidence interval of confidence coefficient  $\alpha$ . Suppose the upper and lower bounds are respectively  $\delta_1$  and  $\delta_2$ . Evaluate the expression for the expected probability of misclassification substituting in turn  $\delta_1$  and  $\delta_2$  for  $\delta$ . The two values thus obtained will enclose the true value of that quantity with probability  $\alpha$ .

Confidence bounds for  $\delta$  can be set up if we remember that

$$N_1(\bar{\mathbf{x}}^{(1)} - \mathbf{y}^{(2)})\Sigma^{-1}(\bar{\mathbf{x}}^{(1)} - \mathbf{y}^{(2)})', N_2(\bar{\mathbf{x}}^{(2)} - \mathbf{y}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \mathbf{y}^{(1)})', \\ (N_1 + N_2)^{-1}N_1N_2(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\Sigma^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$$

have noncentral chisquare distributions with  $p$  degrees of freedom and noncentrality parameters equal respectively to  $N_1\frac{1}{2}\delta^2$ ,  $N_2\frac{1}{2}\delta^2$ , and  $(N_1 + N_2)^{-1}N_1N_2\frac{1}{2}\delta^2$  and that  $[p(N_1 + N_2)(N_1 + N_2 - 2)]^{-1}N_1N_2(N_1 + N_2 - p - 1)(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$  has the noncentral  $F$ -distribution with degrees of freedom  $p$  and  $N_1 + N_2 - p - 1$  and noncentrality equal to  $(N_1 + N_2)^{-1}N_1N_2\frac{1}{2}\delta^2$ .

For the sake of definiteness, let us suppose that we are dealing with a situation where  $\Sigma$  is known and  $\mathbf{y}^{(1)}$  and  $\mathbf{y}^{(2)}$  are unknown. Let  $v$  be a noncentral chisquare variable with  $p$  degrees of freedom and noncentrality equal to  $(N_1 + N_2)^{-1}N_1N_2\frac{1}{2}\delta^2$ . Set

$$(88) \quad F_\delta(z) = \Pr(v < z).$$

As  $\delta_1$  we can take<sup>6</sup> the least upper bound of the set of all numbers  $\delta$  satisfying the

<sup>6</sup> Marakathavalli [12] should also be consulted. She discusses how unbiased critical regions can be set up for testing hypotheses specifying the value of the noncentrality parameter. Inversion will give an unbiased confidence interval. Methods of approximately evaluating the probability integral of the noncentral chisquare density developed in [1] and [13] will be required. Tables given in [6] and [12] will be found useful.

condition

$$(89) \quad F_2([N_1 + N_2]^{-1} N_1 N_2 (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}) \Sigma^{-1} (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})') \geq \frac{1}{2}(1 + \alpha)$$

and as  $\delta_2$  we can take the greatest lower bound of the set of all numbers  $\delta$  satisfying the condition

$$(90) \quad F_2([N_1 + N_2]^{-1} N_1 N_2 (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}) \Sigma^{-1} (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})') \leq \frac{1}{2}(1 - \alpha).$$

**12. About a more general classification procedure.**<sup>7</sup> There are situations where the two kinds of errors are not of equal importance. In some cases it may even be possible to determine the different losses consequent on each type of mistake. Suppose  $c_{12}$  is the loss incurred in assigning an individual from  $P^{(1)}$  to  $P^{(2)}$  and  $c_{21}$  is the loss incurred in assigning an individual from  $P^{(2)}$  to  $P^{(1)}$ . Suppose further that the *a priori* probability of an individual coming from  $P^{(k)}$  is  $\pi^{(k)}$ . Then the procedure with minimum expected loss is that of assigning an individual with measurements  $\mathbf{x}$  to  $P^{(1)}$  or  $P^{(2)}$  according as

$$(91) \quad (\mathbf{y}^{(2)} - \mathbf{y}^{(1)}) \Sigma^{-1} \mathbf{x}' \leq \frac{1}{2} (\mathbf{y}^{(2)} - \mathbf{y}^{(1)}) \Sigma^{-1} (\mathbf{y}^{(2)} + \mathbf{y}^{(1)})' + c$$

where

$$(92) \quad c = \log_e \frac{\pi^{(1)} c_{12}}{\pi^{(2)} c_{21}}$$

[3], p. 134). The procedure we considered earlier is a special case of this more general procedure. It corresponds to the case  $c = 0$ . A sufficient condition for  $c$  to be zero is that  $c_{12} = c_{21}$  and  $\pi^{(1)} = \pi^{(2)}$ .

The procedure mentioned above can be carried out only if all the parameters  $\mathbf{y}^{(1)}$ ,  $\mathbf{y}^{(2)}$  and  $\Sigma$  are known. If such is not the case we may assign the individual to  $P^{(1)}$  or  $P^{(2)}$  according as

$$(93) \quad (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}) \mathbf{S}^{-1} \mathbf{x}' \leq \frac{1}{2} (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}) \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(2)} + \bar{\mathbf{x}}^{(1)})' + c.$$

The sampling fluctuations of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})$  and  $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})$ , the conditional probabilities of the two kinds of errors, are again of interest. We shall briefly discuss the case of  $\Sigma$  known and give indications of the changes to be made in some of the earlier formulae, leaving a fuller discussion of the procedure to another occasion.

The distribution of  $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  is given by the equation

$$(94) \quad \Pr(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) < z) = \Pr(u_1 > -G^{-1}(z))$$

where

$$(95) \quad u_1 = \frac{1}{2} [\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}) \Sigma^{-1} (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{\frac{1}{2}} + [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}) \Sigma^{-1} (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{-\frac{1}{2}} [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}) \Sigma^{-1} (\bar{\mathbf{x}}^{(1)} - \mathbf{y}^{(1)})' + c].$$

<sup>7</sup> The results of earlier sections can be generalized in other ways, which we hope to indicate in a later communication.

This equation shows that it suffices to obtain the distribution of  $u_1$ . The density function  $h(u_1)$  of  $u_1$  is

$$(96) \quad \iint \left( \frac{N_1 + N_2}{2\pi} \right)^{\frac{1}{2}} \exp \left[ -\frac{1}{2} (N_1 + N_2) \left\{ u_1 - \frac{1}{2} \cdot \frac{N_1 - N_2}{N_1 + N_2} t_3^{\frac{1}{2}} \right. \right. \\ \left. \left. - \left( \frac{N_2 t_4}{N_1 + N_2} + c \right) t_3^{\frac{1}{2}} \right\}^2 \right] f(t_3, t_4) dt_3 dt_4$$

where  $f(t_3, t_4)$  is to be taken from equation (56).

If  $N_1, N_2$  and  $n$  are large, we have, corresponding to equation (71),

$$(97) \quad \Pr(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) < z) \approx G(2\delta^2[(\delta^2 + 2c)^2 N_1^{-1} \\ + (\delta^2 - 2c)^2 N_2^{-1} + 2(2c\delta)^2 n^{-1}]^{-\frac{1}{2}} [G^{-1}(z) + c\delta^{-1} + \frac{1}{2}\delta]).$$

If more information is available, the results become simpler. Thus if  $\mathbf{u}^{(1)}$  is also known with  $\Sigma$ , the distribution function of  $e_{12}(\bar{\mathbf{x}}^{(2)})$  is given by the formula

$$(98) \quad \Pr(e_{12}(\bar{\mathbf{x}}^{(2)}) < z) = \begin{cases} \Pr(v < A_1(z)) + \Pr(v > A_2(z)) & \text{if } z < 1 - G([2c]^{\frac{1}{2}}), \quad (c \geq 0) \\ 1 & \text{if } z \geq 1 - G([2c]^{\frac{1}{2}}), \end{cases}$$

where  $v$  is a random variable having the noncentral chisquare distribution with  $p$  degrees of freedom and noncentrality  $N_2 \frac{1}{2} \delta^2$  and

$$(99) \quad A_1(z) = N_2 [G^{-1}(z) + \{[G^{-1}(z)]^2 - 2c\}^{\frac{1}{2}}]^2$$

and

$$(100) \quad A_2(z) = N_2 [G^{-1}(z) - \{[G^{-1}(z)]^2 - 2c\}^{\frac{1}{2}}]^2.$$

Similarly we have, for the distribution function of  $e_{21}(\bar{\mathbf{x}}^{(2)})$ , the equation

$$(101) \quad \Pr(e_{21}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr(w < G^{-1}(z))$$

where the random variable  $w$  has the density function  $h(w)$  defined below.

$$(102) \quad h(w) = \begin{cases} C e^{-N_2 \frac{1}{2} (-2c + \delta^2)} \left[ \int_{m_1(w)}^{m_2(w)} + \int_{m_3(w)}^{m_4(w)} \right] [1 - \delta^{-2} (w - \frac{1}{2} u \\ - c u^{-1})^2]^{\frac{1}{2}(p-3)} u^{p-1} e^{-N_2 u w} du & \text{if } w \geq \delta + (2c)^{\frac{1}{2}} \\ C e^{-N_2 \frac{1}{2} (-2c + \delta^2)} \int_{m_1(w)}^{m_4(w)} [1 - \delta^{-2} (w - \frac{1}{2} u - c u^{-1})^2]^{\frac{1}{2}(p-3)} & (c \geq 0) \\ u^{p-1} e^{-N_2 u w} du & \text{if } (2c)^{\frac{1}{2}} - \delta \leq w < (2c)^{\frac{1}{2}} + \delta \\ 0 & \text{if } w < (2c)^{\frac{1}{2}} - \delta \end{cases}$$

Here



$$\begin{aligned}
 m_1(w) &= w + \delta - [(w + \delta)^2 - 2c]^{\frac{1}{2}}, \\
 m_2(w) &= w - \delta - [(w - \delta)^2 - 2c]^{\frac{1}{2}}, \\
 m_3(w) &= w - \delta + [(w - \delta)^2 - 2c]^{\frac{1}{2}}, \\
 m_4(w) &= w + \delta + [(w + \delta)^2 - 2c]^{\frac{1}{2}},
 \end{aligned}
 \tag{103}$$

and  $C$  is the  $C$  of equation (42).

Observe that with probability one,

$$e_{12}(\bar{\mathbf{x}}^{(2)}) \leq 1 - G([2c]^{\frac{1}{2}}) \quad (c \geq 0) \tag{104}$$

and

$$e_{21}(\bar{\mathbf{x}}^{(2)}) \geq G([2c]^{\frac{1}{2}} - \delta) \quad (c \geq 0). \tag{105}$$

If  $c < 0$ , we have, for the distribution function of  $e_{12}(\bar{\mathbf{x}}^{(2)})$ , the equation

$$\Pr(e_{12}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr(v > A_2(z)) \tag{106}$$

instead of equation (98) and, for the density function of  $w$ , the equation

$$\begin{aligned}
 h(w) &= Ce^{-N_2 \frac{1}{2}(-2c + \delta^2)} \int_{m_3(w)}^{m_4(w)} [1 - \delta^{-2}(w - \frac{1}{2}u \\
 &\quad - cu^{-1})^2]^{\frac{1}{2}(p-3)} u^{p-1} e^{-N_2 w u} du
 \end{aligned}
 \tag{107}$$

instead of equation (102).

Note that statements corresponding to equations (104) and (105) cannot be made if  $c < 0$ .

For  $Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$  we have the following result:

$$Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = \int_d^\infty f_1(z) dz \tag{108}$$

where  $f_1(z)$  is the function defined by equation (4.9) of [10] and

$$d = 4N_1N_2(N_1 + N_2)^{-1}(N_1 + N_2 + 4N_1N_2)^{-1}c. \tag{109}$$

As the expression for  $f_1(z)$  is complicated we shall not reproduce it here.

**Acknowledgment.** The author wishes to express his gratitude to Dr. C. R. Rao for his kind encouragement.

# REFERENCES

- [1] S. H. ABDEL-ATY, "Approximate formula for the percentage points and the probability integral of the noncentral  $\chi^2$  distribution," *Biometrika*, Vol. 41 (1954), pp. 538-540.
- [2] T. W. ANDERSON, "Classification by multivariate analysis," *Psychometrika*, Vol. 16 (1951), pp. 31-50.
- [3] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

- [4] M. M. BARNARD, "The secular variations of skull characters in four series of Egyptian skulls," *Ann. Eugen.*, Vol. 6 (1935), pp. 352-371.
- [5] R. C. BOSE AND S. N. ROY, "The exact distribution of the studentized  $D^2$ -statistic," *Sankhyā*, Vol. 4 (1938), pp. 19-38.
- [6] R. A. FISHER, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc., Ser. A*, Vol. 121, (1928), pp. 654-673.
- [7] R. A. FISHER, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, Vol. 7 (1936), pp. 179-188.
- [8] HARMAN LEON HARTER, "On the distribution of Wald's classification statistic," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 58-67.
- [9] S. JOHN, "The distribution of Wald's classification statistic when the dispersion matrix is known," *Sankhyā*, Vol. 21 (1959), pp. 371-376.
- [10] S. JOHN, "On some classification statistics," *Sankhyā*, Vol. 22 (1960), pp. 309-317.
- [11] CARL F. KOSSACK, "Some techniques for simple classification," *Proceedings of the Berkeley Symposium in Statistics and Probability*, pp. 345-352, University of California Press, Berkeley, 1949.
- [12] N. MARAKATHAVALLI, "Unbiased test for a specified value of a parameter in the non-central  $F$ -distribution," *Sankhyā*, Vol. 15 (1955), pp. 321-330.
- [13] P. B. PATNAIK, "The noncentral  $\chi^2$  and  $F$  distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202-232.
- [14] E. S. PEARSON AND H. O. HARTLEY, *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press, 1956.
- [15] National Bureau of Standards, *Tables of the Bivariate Normal Distribution and Related Functions*, Washington, D. C., 1959.
- [16] KARL PEARSON, *Tables of the Incomplete Beta-Function*, Cambridge, Cambridge University Press, 1934.
- [17] C. RADHAKRISHNA RAO, "On a general theory of discrimination when the information on alternate hypotheses is based on samples," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 651-669.
- [18] C. RADHAKRISHNA RAO, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.
- [19] ROSEDDITH SITGREAVES, "On the distribution of two random matrices used in classification procedures," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 263-270.
- [20] C. A. B. SMITH, "Some examples of discrimination," *Ann. Eugen.*, Vol. 13 (1947), pp. 272-282.
- [21] ERLING SYVERDRUP, "Derivation of the Wishart distribution of the second order sample moments by straightforward integration of a multiple integral," *Skandinavisk Aktuarietidskrift*, Vol. 30 (1947), pp. 151-166.
- [22] A. WALD, "On a statistical problem arising in the classification of an individual into one of two groups," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 145-163.
- [23] B. L. WELCH, "Note on discriminant functions," *Biometrika*, Vol. 31 (1939), pp. 218-220.

# ON THE MONOTONIC CHARACTER OF THE POWER FUNCTIONS OF TWO MULTIVARIATE TESTS<sup>1</sup>

BY S. N. ROY AND W. F. MIKHAIL

*University of North Carolina*

**1. Summary and introduction.** The largest characteristic root has been proposed in [2] as a test statistic in (i) the multivariate analysis of variance test, and (ii) testing that two sets of variates are independent. In this paper it is shown that, in each case, the power function is a monotonically increasing function of each non-centrality parameter, separately. This property was stated in [2] without proof. This provides a stronger result than would be obtained by any direct use of Anderson's Theorem [1] which implies that the power function increases when all the roots are simultaneously increased in the same ratio. The proof of the monotonicity property for the multivariate analysis of variance is given in Section 3, and in Section 4 it is shown how the proof is modified for testing independence between two sets of variates.

**2. Preliminaries for the multivariate analysis of variance situation.** Let  $u, s$  and  $n$  denote respectively the "effective" number of variates, the degrees of freedom of the hypothesis and the degrees of freedom for the error and let  $t = \min(u, s)$ . Then, with  $\mathbf{X} \equiv [x_{ij}]: u \times s$  and  $\mathbf{Y} \equiv [y_{ij}]: u \times n$ , the canonical form for the d.f. in the multivariate analysis of variance model was obtained in [2, p. 86] to be

$$(2.1) \quad \begin{aligned} & [1/(2\pi)^{\frac{1}{2}u(s+n)}] \\ & \cdot \exp \left[ -\frac{1}{2} \left\{ \sum_{i=1}^u \sum_{j=1}^n y_{ij}^2 + \sum_{i=1}^t (x_{ii} - \theta_i)^2 + \sum_{i=t+1}^u x_{ii}^2 + \sum_{i=1}^u \sum_{j=t+1}^s x_{ij}^2 \right\} \right] \\ & \cdot \prod_{i=1}^u \prod_{j=1}^s dx_{ij} \prod_{i=1}^u \prod_{j=1}^n dy_{ij} \equiv Q \, d\mathbf{X} \, d\mathbf{Y}, \end{aligned}$$

and the acceptance region of size  $1 - \alpha$  for the linear hypothesis  $H_0$  of analysis of variance, i.e., for the case  $\theta_1 = \theta_2 = \dots = \theta_t = 0$ , can be expressed as

$$(2.2) \quad \mathcal{D} = \{ \mathbf{X}, \mathbf{Y} : c_M[(\mathbf{X}\mathbf{X}')(\mathbf{Y}\mathbf{Y}')^{-1}] \leq \mu \},$$

where  $\mu$  is given by

$$(2.3) \quad P[\mathbf{X}, \mathbf{Y} \in \mathcal{D} \mid \theta'_is = 0] = 1 - \alpha,$$

and where the  $\theta'_i$ 's are the non-centrality parameters defined in [2, pp. 85-86],

Received September 1, 1959; revised April 5, 1961.

<sup>1</sup> This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

and  $c_M(\mathbf{A})$  denotes the largest characteristic root of the (square) matrix  $\mathbf{A}$ . The problem now is to prove that the integral,  $P[\mathbf{X}, \mathbf{Y} \in \mathcal{D}] = \int_{\mathcal{D}} Q d\mathbf{X} d\mathbf{Y}$  is a monotonically decreasing function of each  $\theta_i^2$ , separately. If we regard the domain  $\mathcal{D}$  as one of dimensionality  $u s + u n$ , where  $u s$  dimensions are associated with  $\mathbf{X}$  and  $u n$  with  $\mathbf{Y}$ , then it is clear that we can rewrite this integral as

$$(2.4) \quad \int_{\mathcal{D}^*} \text{const.} \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^u \sum_{j=1}^n y_{ij}^2 + \sum_{i=1}^u \sum_{j=1}^s x_{ij}^2 \right) \right] d\mathbf{X} d\mathbf{Y} \\ = \int_{\mathcal{D}^*} Q^* d\mathbf{X} d\mathbf{Y}, \quad (\text{say}),$$

where  $\mathcal{D}^*$  is merely the domain  $\mathcal{D}$  translated by  $\theta_i$  along  $x_{ii}$ , that is, along the  $ii$ th axis (with  $i = 1, 2, \dots, t$ ). Notice that if, in the integral (2.4), we replace the domain  $\mathcal{D}^*$  by  $\mathcal{D}$ , the integral over the new domain becomes equal to  $1 - \alpha$ , where  $\alpha$  is the probability of the first kind of error. Let

$$(2.5) \quad \mathbf{Y}\mathbf{Y}' = (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1},$$

where  $\tilde{\mathbf{V}}$  is a  $u \times u$  triangular matrix with zeros above the main diagonal. Observe that

$$c_M[(\mathbf{X}\mathbf{X}')(\mathbf{Y}\mathbf{Y}')^{-1}] = c_M[(\mathbf{X}\mathbf{X}')(\tilde{\mathbf{V}}'\tilde{\mathbf{V}})] = c_M[(\tilde{\mathbf{V}}\mathbf{X})(\tilde{\mathbf{V}}\mathbf{X})'],$$

and rewrite (2.2), that is, the domain  $\mathcal{D}$  as

$$(2.6) \quad \mathcal{D} = \{\mathbf{Y}, \mathbf{X} : c_M[\tilde{\mathbf{V}}\mathbf{X}(\tilde{\mathbf{V}}\mathbf{X})'] \leq \mu\}.$$

Notice that  $\tilde{\mathbf{V}}$  is a function of  $\mathbf{Y}$  given by (2.5).

The problem now can be rephrased in the following way. How does the integral of  $Q^*$  given by (2.4) over the domain  $\mathcal{D}$  given by (2.6) change under successive translations of  $\theta_1$  along  $x_{11}$ , of  $\theta_2$  along  $x_{22}$ ,  $\dots$ ,  $\theta_t$  along  $x_{tt}$ ? It is clear that the successive changes are cumulative. It will be also seen from the mechanics of the demonstration that if we can prove that the integral decreases for the first shift of  $\mathcal{D}$ , namely, by  $\theta_1$  along  $x_{11}$ , then the general theorem itself will be proved.

**3. Proof of the monotonicity property for the multivariate analysis of variance situation.** The proof is developed in three main steps discussed in the following subsections.

**3.1 The proof for the univariate case.** In this case,  $u = 1$  and we can drop the first subscript in  $\mathbf{X}, \mathbf{Y}$ . The domain  $\mathcal{D}$  of (2.6) now takes on the form

$$(3.1.1) \quad \mathcal{D} = \left\{ \mathbf{X}, \mathbf{Y} : \sum_{j=1}^s x_j^2 \leq \mu \sum_{j=1}^n y_j^2 \right\},$$

and the integral of  $Q$  over this domain takes the final form

$$(3.1.2) \quad \int_{\mathcal{D}^*} \exp \left[ -\frac{1}{2} \left( \sum_{j=1}^n y_j^2 + \sum_{j=1}^s x_j^2 \right) \right] \prod_{j=1}^n dy_j \prod_{j=1}^s dx_j.$$

Notice that now  $\mathfrak{D}^*$  is just a shift of  $\mathfrak{D}$  along  $x_1$  by  $\theta$ . It is evident from the form of (3.1.2) that the integral (3.1.2) decreases under this shift if, for any given set of  $y_j$ 's for  $j = 1, 2, \dots, n$ , and  $x_j$ 's, for  $j = 2, 3, \dots, s$ ,

$$(3.1.3) \quad \int_{-a+\theta}^{a+\theta} \exp \left[ -\frac{1}{2} x_1^2 \right] dx_1 < \int_{-a}^a \exp \left[ -\frac{1}{2} x_1^2 \right] dx_1,$$

where  $a = +[\mu \sum_{j=1}^s y_j^2 - \sum_{j=2}^s x_j^2]^{\frac{1}{2}}$ ; it is clear that it doesn't matter whether we take  $\theta$  to be positive or negative. It is easy to verify this and also an even more general result, namely that

$$(3.1.4) \quad \int_{-a+\lambda}^{a+\lambda} \phi(x) dx \leq \int_{-a}^a \phi(x) dx,$$

for all real  $\lambda$  and  $a > 0$ , where  $\phi(x)$  is a continuous function of  $x$ , symmetric about 0 and monotonically decreasing with  $|x|$ . It is also clear that the left side of (3.1.3) monotonically decreases with  $|\theta|$ .

3.2. *The nature of the multivariate domain* (2.6). We now characterize  $\mathfrak{D}$  as a domain in  $(x_{11}, x_{21}, \dots, x_{u1})$  for fixed values of  $\bar{V}$ ,  $\mu$  and  $\mathbf{X}$  (excluding the first column). Toward this end, put  $\mathbf{X}^* = \bar{V}\mathbf{X}$  and observe that, if  $\nu$  is any characteristic root of  $\mathbf{X}^*\mathbf{X}^{*'} = \mathbf{S}^*$ , then

$$(3.2.1) \quad |\mathbf{S}^* - \nu \mathbf{I}| = 0 = |\mathbf{S}^*| - \nu \text{tr}_{p-1} + \nu^2 \text{tr}_{p-2} - \dots + (-1)^u \nu^u,$$

where  $\text{tr}_j$  is the trace of the  $j$ th order, or in other words, the sum of the  $j$ th rowed principal minor determinants of  $\mathbf{S}^*$ . But, given  $x_{ij}^*$  (for  $i = 1, 2, \dots, u$ ;  $j = 2, \dots, s$ ),  $|\mathbf{S}^*| = |\mathbf{X}^*\mathbf{X}^{*'}|$  is a homogeneous quadratic function of  $(x_{11}^*, \dots, x_{u1}^*) +$  a constant which is really a function of the other  $x_{ij}^*$ 's just mentioned. The coefficients of the quadratic function are also each a polynomial function of  $x_{ij}^*$ 's (for  $i = 1, 2, \dots, u$ ;  $j = 2, 3, \dots, s$ ). Likewise, if we take any  $q$ -rowed principal minor determinant of  $[s_{ij}^*]$ , say the one with rows and columns numbered, 1, 2,  $\dots$ ,  $q$ , then that determinant is

$$\begin{vmatrix} x_{11}^* & \dots & x_{1s}^* \\ \vdots & \ddots & \vdots \\ x_{q1}^* & \dots & x_{qs}^* \end{vmatrix} \begin{vmatrix} x_{11}^* & \dots & x_{q1}^* \\ \vdots & \ddots & \vdots \\ x_{1s}^* & \dots & x_{qs}^* \end{vmatrix},$$

which, given the other  $x_{ij}^*$ 's, is a homogeneous quadratic function of  $(x_{11}^*, \dots, x_{q1}^*)$  (in which the coefficients are polynomials in the other  $x_{ij}^*$ 's) + a constant which is really a function of the other  $x_{ij}^*$ 's. Thus, given  $\nu$  and the other  $x_{ij}^*$ 's, the equation (3.2.1) in  $\nu$  yields a homogeneous quadric surface in  $x_{11}^*, \dots, x_{u1}^*$ . Now recall from (2.5) that, given  $y_{ij}$ 's, that is  $\bar{V}$ , the  $(x_{11}^*, \dots, x_{u1}^*)$  are linear functions of  $(x_{11}, \dots, x_{u1})$  and likewise  $(x_{i2}^*, \dots, x_{is}^*)$  are linear functions of  $(x_{1j}, \dots, x_{uj})$ , for  $j = 2, \dots, s$ . Thus, given  $\nu$  and

$$x_{ij} \text{ (for } i = 1, 2, \dots, u; j = 2, \dots, s),$$

the equation (3.2.1) yields a homogeneous quadric surface in  $(x_{11}, \dots, x_{u1})$  in which the coefficients and the constant term are all functions of  $\nu$ ,  $\mathbf{Y}$  and the other  $x_{ij}$ 's already referred to. This is for any characteristic root  $\nu$ .

Let us now rewrite (2.2) in the (almost everywhere) equivalent form

$$(3.2.2) \quad \mathfrak{D} = \left\{ \mathbf{X}, \mathbf{Y} : \sup_a \frac{(x_{11} + a_2 x_{21} + \cdots + a_u x_{u1})^2 + \cdots + (x_{1s} + a_2 x_{2s} + \cdots + a_u x_{us})^2}{(y_{11} + a_2 y_{21} + \cdots + a_u y_{u1})^2 + \cdots + (y_{1s} + a_2 y_{2s} + \cdots + a_u y_{us})^2} \leq \mu \right\},$$

where  $\mathbf{a}' = (a_2, \dots, a_u)$ . Now, given  $\mu$ ,  $\mathbf{Y}$  and  $x_{ij}$ 's (for  $i = 1, \dots, u; j = 2, \dots, s$ ), (3.2.2) represents the domain of  $(x_{11}, \dots, x_{u1})$  in an  $u$ -dimensional Euclidean space, the boundary being given by the surface defined by the equality sign. An equivalent form of the same surface is the homogeneous quadric associated with (3.2.1) after  $\nu$  is replaced by  $\mu$ . Next, it is easy to check from the definition of  $\mathfrak{D}$  and the manner in which the vector  $(x_{11}, \dots, x_{u1})$  occurs in it that (3.2.2) implies the following:

(i)  $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}$  implies that  $((c\mathbf{x}_1, \mathbf{X}_2), \mathbf{Y}) \in \mathfrak{D}$  for  $0 \leq c \leq 1$ , where  $\mathbf{X}$  is decomposed into  $(\mathbf{x}_1, \mathbf{X}_2)$  such that  $\mathbf{x}'_1 = (x_{11}, \dots, x_{u1})$  and  $\mathbf{X}_2$  is a matrix with the other elements of  $\mathbf{X}$ .

(ii) Any straight line passing through the origin  $(0, \dots, 0)$  has an intersection with the domain, of finite length.

Thus, given  $\mu$ ,  $\mathbf{Y}$  and the other  $x_{ij}$ 's (already described), (2.2) or (2.6) can be regarded as a domain of  $(x_{11}, \dots, x_{u1})$  which is the interior of a  $u$ -dimensional ellipsoid whose boundary is given by (3.2.1) after  $\mu$  is substituted for  $\nu$ . It is well known that there is an orthogonal transformation by which the ellipsoid can be referred to principal axes, or in other words, the transformed equation to the surface becomes free from the product terms in the transformed variables and involves only the square terms with positive coefficients. Let  $\mathbf{x}'_1 = [x_{11}, \dots, x_{u1}]$  and

$$(3.2.3) \quad \mathbf{z} = \mathbf{L}\mathbf{x}_1,$$

where  $\mathbf{L}: u \times u$  is an orthogonal matrix that transforms the ellipsoid into principal axes. This  $\mathbf{L}$  can be determined and the rows of  $\mathbf{L}$ , say  $(l_{i1}, \dots, l_{iu})$ ,  $i = 1, 2, \dots, u$ , are the direction cosines of the different principal axes. Note that  $\mathbf{z}'\mathbf{z} = \mathbf{x}'_1\mathbf{x}_1$ . It would be useful to rewrite (2.4), after substitution of  $\mathfrak{D}$  for  $\mathfrak{D}^*$  and omission of the constant, in the form

$$(3.2.4) \quad \int_{\mathfrak{D}} \exp \left[ -\frac{1}{2} \left\{ \sum_{i=1}^u \sum_{j=1}^n y_{ij}^2 + \sum_{i=1}^u \sum_{j=2}^s x_{ij}^2 + \sum_{i=1}^u z_i^2 \right\} \right] d\mathbf{Y} \prod_{i=1}^u \prod_{j=2}^s dx_{ij} \prod_{i=1}^u dz_i,$$

where, given  $\mu$ ,  $\mathbf{Y}$  and the  $x_{ij}$ 's, the domain  $\mathfrak{D}$ , as a domain in  $(z_1, \dots, z_u)$ , forms the interior of an ellipsoid referred to principal axes (that is, in a form which is free from the product terms of  $z$ 's and involves only the square terms with positive coefficients). In other words,  $\mathfrak{D}$  is symmetric about the origin in each  $z_i$  separately. A displacement  $\theta_1$  along the direction of  $x_{11}$  might be regarded as the resultant of a displacement  $l_{11}\theta_1$  along  $z_1$ , that is, along the direction with cosines  $(l_{11}, l_{12}, \dots, l_{1u})$ , a displacement  $l_{21}\theta_1$  along  $z_2$ , that is, along the direc-



tion with cosines ( $l_{21}, l_{22}, \dots, l_{2u}$ ), and so on, and finally a displacement  $l_{u1}\theta_1$  along  $z_u$ , that is, along the direction with cosines ( $l_{u1}, \dots, l_{uu}$ ). It should be remembered that these  $l_{ij}$ 's are functions of  $\mu, \mathbf{Y}$  and the  $x_{ij}$ 's of (3.2.4).

3.3 *The final step in the proof of the monotonicity property.* Looking at (3.2.4) and using (3.1.4) we observe that a displacement of  $\mathcal{D}$  by  $l_{11}\theta_1$  along  $z_1$  will decrease the integral under (3.2.4), because, for any given set  $\mu, \mathbf{Y}, x_{ij}$ 's and  $z_2, z_3, \dots, z_u$ ,

$$(3.3.1) \quad \int_{-a+l_{11}\theta_1}^{a+l_{11}\theta_1} \exp\left[-\frac{1}{2}z_1^2\right] dz_1 < \int_a^a \exp\left[-\frac{1}{2}z_1^2\right] dz_1,$$

where  $a$  and  $l_{11}\theta_1$ , without any loss of generality, can be assumed to be positive. Recall that  $a$  is a function of  $\mu, \mathbf{Y}, x_{ij}$ 's and  $z_2, \dots, z_u$ . Using the same argument for successive displacements by  $l_{21}\theta_1$  along  $z_2$ , by  $l_{31}\theta_1$  along  $z_3$ , and so on, and finally by  $l_{u1}\theta_1$  along  $z_u$  we have successive decreases of the integral. In other words, the resultant displacement which is along  $x_{11}$  and by  $\theta_1$  decreases the integral. At this point we go back to the integral over  $\mathcal{D}$  of  $Q^*$ , forget about the  $z_i$ 's, use the result just stated about a displacement by  $\theta_1$  along  $x_{11}$ , apply successive displacements by  $\theta_2$  along  $x_{22}$ ,  $\theta_3$  along  $x_{33}$  and so on, and finally  $\theta_t$  along  $x_{tt}$  and eventually obtain an integral over the displaced domain  $\mathcal{D}^*$  which is less than the one over the original domain  $\mathcal{D}$ . It is also clear from the mechanics of the proof that the integral over  $\mathcal{D}^*$  decreases as each  $|\theta_i|$ ,  $i = 1, 2, \dots, t$ , increases separately. This proves the monotonicity property.

4. **The case of the test for independence between two sets of variates.** With a  $(p+q)$  set ( $p \leq q$ ) of variables let us assume, for a sample of size  $n+1$  ( $> p+q$ ), the canonical distribution law ([2], p. 68)

$$(4.1) \quad \left[ 1/(2\pi)^{\frac{1}{2}(p+q)n} \prod_{i=1}^p (1 - \rho_i^2)^{n/2} \right] \cdot \exp \left[ -\frac{1}{2} \left\{ \sum_{i=1}^p \frac{1}{1 - \rho_i^2} (x_{ij}^2 + y_{ij}^2 - 2\rho_i x_{ij} y_{ij}) + \sum_{i=p+1}^q \sum_{j=1}^n y_{ij}^2 \right\} \right] \cdot \prod_{i=1}^p \prod_{j=1}^n dx_{ij} \prod_{i=1}^q \prod_{j=1}^n dy_{ij},$$

where  $\rho_i$ 's are the population canonical correlation coefficients. The hypothesis of independence  $H_0$  is equivalent to the hypothesis that  $\rho_i$ 's = 0; the acceptance region (of size  $1 - \alpha$ ) for  $H_0$  is

$$(4.2) \quad \mathcal{D} = \{\mathbf{X}, \mathbf{Y}; c_M[(\mathbf{X}\mathbf{X}')^{-1}(\mathbf{X}\mathbf{Y}')(\mathbf{Y}\mathbf{Y}')^{-1}(\mathbf{Y}\mathbf{X}')] \leq \mu\},$$

where  $\mu$  is given by

$$P[\mathbf{X}, \mathbf{Y} \in \mathcal{D} \mid H_0] = 1 - \alpha.$$

The monotonicity in this case is proved in exactly the same way as in the previous case. For this purpose we rewrite the  $\mathcal{D}$  of (4.2) as

$$(4.3) \quad \mathfrak{D} = \left\{ \mathbf{X}, \mathbf{Y}: c_M[(\mathbf{U}\mathbf{U}')(\mathbf{V}\mathbf{V}')^{-1}] \leq \frac{\mu}{1-\mu}, \text{ i.e., } \leq \mu^* \text{ (say)} \right\},$$

and the d.f. of (4.1) as

$$(4.4) \quad \text{const. exp} \left[ -\frac{1}{2} \left( \sum_{i=1}^p \sum_{j=1}^q (u_{ij} - \gamma_{ij} t_{ij})^2 + \sum_{i=1}^q \sum_{j=1}^i t_{ij}^2 + \sum_{i=1}^p \sum_{j=q+1}^n v_{ij}^2 \right) \right] \\ \cdot d\mathbf{U} d\mathbf{V} \prod_{i=1}^q t_{ii}^{n-i} d\tilde{\mathbf{T}},$$

where

$$(4.5) \quad \gamma_{ij} = \rho_i / (1 - \rho_i^2)^{\frac{1}{2}} = \theta_i \text{ (say)}, \\ (\text{for } j = 1, 2, \dots, i; i = 1, 2, \dots, p), \text{ and } \gamma_{ij} = 0,$$

otherwise, and where  $\tilde{\mathbf{T}}: q \times q$ ,  $\mathbf{U}: p \times q$ , and  $\mathbf{V}: p \times (n - q)$  are related to  $(\mathbf{X}: p \times n, \mathbf{Y}: q \times n)$  in the following way:

$$(4.6) \quad \mathbf{Y} = \tilde{\mathbf{T}}\mathbf{L},$$

where  $\mathbf{L}: q \times n$  is orthonormal and  $\tilde{\mathbf{T}}$  is lower triangular.  $\mathbf{M}: (n - q) \times n$  is an orthogonal completion of  $\mathbf{L}$ ,  $\mathbf{D}_{a_i}; p \times p$  stands for a diagonal matrix with diagonal elements  $a_1, a_2, \dots, a_p$ , and  $\mathbf{U}$  and  $\mathbf{V}$  are given by

$$(4.7) \quad \mathbf{U} = \mathbf{D}_{(1-\rho_i^2)}^{-\frac{1}{2}} \mathbf{X}\mathbf{L}', \quad \mathbf{V} = \mathbf{D}_{(1-\rho_i^2)}^{-\frac{1}{2}} \mathbf{X}\mathbf{M}'.$$

In the transformation from (4.1) to (4.4),  $\mathbf{M}$  does not occur explicitly,  $\mathbf{L}$  does, but is easily integrated out as in [2, pp. 196-197].

The probability of the second kind of error is given by integrating (4.4) over the domain (4.3). It is easy to see that, aside from the positive constant factor, this is equivalent to

$$(4.8) \quad \int_{\mathfrak{D}^*} \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^q \sum_{j=1}^i t_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^q u_{ij}^2 + \sum_{i=1}^p \sum_{j=q+1}^n v_{ij}^2 \right) \right] d\mathbf{U} d\mathbf{V} \prod_{i=1}^q t_{ii}^{n-i} d\tilde{\mathbf{T}}$$

where, for any given set of  $\tilde{\mathbf{T}}$  and  $\mathbf{V}$ ,  $\mathfrak{D}^*$  is just  $\mathfrak{D}$  displaced by  $\theta_1 t_{11}$  along  $u_{11}$ , by  $\theta_2 t_{21}$  along  $u_{21}$  and  $\theta_2 t_{22}$  along  $u_{22}$ , and so on, and finally by  $\theta_p t_{p1}$  along  $u_{p1}$ ,  $\theta_p t_{p2}$  along  $u_{p2}$ ,  $\dots$ ,  $\theta_p t_{pp}$  along  $u_{pp}$ . Notice that when  $H_0$  is true, that is, when  $\theta_i = 0$ , we should have  $\mathfrak{D}^*$  replaced by  $\mathfrak{D}$  in the integral (4.8). Using the same kind of argument as in Section 3 it follows that, for any given  $\tilde{\mathbf{T}}$ , the partial integral over  $\mathbf{U}$  and  $\mathbf{V}$  decreases as  $\mathfrak{D}$  is displaced by  $\theta_i t_{i1}$  along  $u_{i1}$ , where  $t_{i1} > 0$ , almost everywhere, and with this displacement of  $\mathfrak{D}$ , it is easy to see that the total integral (if we now integrate over  $\tilde{\mathbf{T}}$ ) will also decrease. From considerations of symmetry, the same result would follow for the other displacements in that the displacement associated with any  $\theta_i^2$  could be represented as a  $\theta_i t_{i1}$  along  $u_{i1}$  under a suitable transformation. Thus (4.15) monotonically decreases as each  $|\theta_i|$ , that is, each  $|\rho_i|$  separately increases.

**Concluding remarks.** The power functions of the  $\lambda$ -criteria for the multivariate linear hypothesis and for the test of independence between two sets of

variates have also somewhat similar monotonicity properties that will be discussed in a subsequent paper.

## REFERENCES

- [1] T. W. ANDERSON, "The integral of a symmetrical unimodal function," *Proc. Amer. Math. Soc.*, Vol. 6 (1955), pp. 170-176.
- [2] S. N. ROY, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.
- [3] S. N. ROY AND R. GNANADESIKAN, "Further contributions to multivariate confidence bounds," *Biometrika*, Vol. 44 (1957), pp. 339-410.
- [4] S. N. ROY AND R. GNANADESIKAN, "Some contributions to ANOVA in one or more dimensions—Parts I and II," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 304-340.

# THE MOMENTS OF ELEMENTARY SYMMETRIC FUNCTIONS OF THE ROOTS OF A MATRIX IN MULTIVARIATE ANALYSIS<sup>1</sup>

BY TITO A. MIJARES

*Harvard University and University of the Philippines<sup>2</sup>*

**1. Introduction and summary.** Pillai and Mijares [7] gave the exact expressions for the first four moments of the sum of  $s$  non-zero roots of a matrix occurring in multivariate normal analysis as studied independently by R. A. Fisher [3], P. L. Hsu [4] and S. N. Roy [9]. In this paper some properties of completely homogeneous symmetric functions and certain determinantal results (Section 2) are used to give an inverse derivation of those moments (Section 4). The method is further extended to the moments in general of elementary symmetric functions (e.s.f.) of the roots of a matrix in multivariate analysis (Section 6) through the use of certain properties of compound matrices (Section 5).

**2. Some results to be used in Sections 4 and 6.** Define the completely homogeneous symmetric function (c.h.s.f.) of the  $p$ th degree in  $k$  arguments by

$$(2.1) \quad \phi_p(x_1, \dots, x_k) = \sum_{P(p)} x_1^{p_1} x_2^{p_2} \dots x_k^{p_k},$$

where  $\sum$  extends over all partitions  $P(p)$  of a non-negative integer  $p = \sum_{i=1}^k p_i$ . Define further  $\phi_0 = 1$  and  $\phi_{p'} = 0$  if  $p' < 0$ .

LEMMA 1.

$$\phi_p(x_1, \dots, x_k) = \phi_p(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) + x_i \phi_{p-1}(x_1, \dots, x_k).$$

PROOF. Partition  $\phi_p(x_1, \dots, x_k)$  into two groups, one group to contain  $x_i$  and the other group not to contain  $x_i$ . Factor out  $x_i$  from the first group and take the sum of the two groups.

LEMMA 2.

$$\begin{aligned} (x_{i+j} - x_i) \phi_{p-1}(x_1, \dots, x_k) \\ = \phi_p(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) \\ - \phi_p(x_1, \dots, x_i, \dots, x_{i+j-1}, x_{i+j+1}, \dots, x_k). \end{aligned}$$

PROOF. Use Lemma 1 separately for the  $x_{i+j}$  and  $x_i$  arguments on

$$\phi_p(x_1, \dots, x_k),$$

take the difference of the two resulting equations, and transpose the term  $(x_{i+j} - x_i) \phi_{p-1}(x_1, \dots, x_k)$  to the left of the equality sign.

Received July 25, 1960; revised July 8, 1961.

<sup>1</sup> Done under a fellowship grant from the Bureau of Technical Assistance Operations, United Nations, New York, at Harvard University.

<sup>2</sup> On leave from the University of the Philippines.

THEOREM. Let  $r_1 \neq r_2 \neq \dots \neq r_k$  be non-negative powers of the  $x$ 's in the successive columns of the  $k$ -order determinant given below and let  $\phi_j$  be the c.h.s.f. in all  $k$  arguments. Then

$$(2.2) \quad |x_{k-i+1}^{r_j}| = D |\phi_{r_j-k+i}|, \quad i, j = 1, \dots, k,$$

where

$$(2.3) \quad D = |x_{k-i+1}^{k-j}|$$

and  $x_{k-i+1}^{r_j}, \phi_{r_j-k+i}, x_{k-i+1}^{k-j}$  are the  $(i, j)$ th elements of the square matrices  $(x_{k-i+1}^{r_j}), (\phi_{r_j-k+i}), (x_{k-i+1}^{k-j})$ , respectively.

PROOF. Perform the following elementary operations on the left determinant of (2.2):  $i$ th row  $- k$ th row,  $i = 1, \dots, k-1$ . Use Lemma 2 for two arguments to factor out  $x_{k-i+1} - x_1$  from the  $i$ th row. The result is

$$(2.4) \quad \prod_{j=2}^k (x_j - x_1) \cdot |\phi_{r_j'-k+i'}(x_1, x_{k-i'+1})|, \quad i', j' = 1, \dots, k,$$

where  $\phi_{r_j'-k+i'}(x_1, x_{k-i'+1})$  is the  $(i', j')$ th element of  $|\phi_{r_j'-k+i'}(x_1, x_{k-i'+1})|$ . Thus, the determinantal expression in (2.4) has c.h.s.f. in two arguments for its elements, except for the elements in the last row which have  $x_1$  only as argument.

Next, repeat the operation on the determinant of (2.4) with  $i$ th row  $-(k-1)$ th row,  $i = 1, \dots, k-2$  and use Lemma 2 for three arguments this time in factoring out  $x_{k-i+1} - x_2$  from the  $i$ th row as the resulting determinant. Repeat the operation until finally we have 1st row  $- 2$ nd row and the same lemma is used but for  $k$  arguments. After factoring out  $x_k - x_{k-1}$  from the last determinant, the expression (2.4) reduces finally to

$$(2.5) \quad \prod_{i>j} (x_i - x_j) \cdot |\phi_{r_j'-k+i'}(x_1, x_2, \dots, x_{k-i'+1})|, \quad i', j' = 1, \dots, k,$$

with the  $i'$ th row of elements  $\phi_{r_j'-k+i'}$  containing arguments  $x_1, x_2, \dots, x_{k-i'+1}$ .

The product  $\prod_{i>j} (x_i - x_j)$  is equal to a determinant of Vandermonde and given directly by (2.3). The determinantal part of (2.5) can be reduced into a determinant in the  $\phi$ 's, with complete arguments  $x_1, \dots, x_k$ , by the repeated application of Lemma 1. Take any element at the intersection of the  $i$ th row,  $i = 2, \dots, k$ , and a given  $j$ th column of the determinant in (2.5), and note that

$$(2.6) \quad \phi_{r_j-k+i}(x_1, \dots, x_{k-i+1}) + x_{k-i+2} \phi_{r_j-k+i-1}(x_1, \dots, x_{k-i+2}) \\ = \phi_{r_j-k+i}(x_1, \dots, x_{k-i+2})$$

after using Lemma 1 for  $k-i+2$  arguments. Hence, perform  $i$ th row  $+ x_{k-i+2} \cdot (i-1)$ th row successively for  $i = k, k-1, \dots, 2$  in the order given and use Lemma 1 on  $k-i+2$  arguments. This increases by one the number of arguments of the  $\phi$ 's in each row. Now, perform  $i$ th row  $+ x_{k-i+3} \cdot (i-1)$ th row for  $i = k, k-1, \dots, 3$  and use Lemma 1 on  $k-i+3$  arguments, etc., until

finally we have  $(k-1)$ th row  $+ x_k \cdot k$ th row with Lemma 1 used for  $k$  arguments. This completes all the arguments in every  $\phi$  of the determinant in (2.5).

**3. The mathematical expectation of the sum of the roots.** The well-known distribution of the  $s$  non-zero roots obtained independently by R. A. Fisher [3], P. L. Hsu [4] and S. N. Roy [9] can be written [7] in the form

$$(3.1) \quad f(\theta_1, \dots, \theta_s) = c(s, m, n) \prod_{i=1}^s \theta_i^m (1 - \theta_i)^n \prod_{i>j} (\theta_i - \theta_j),$$

$$0 < \theta_1 \leq \dots \leq \theta_s < 1,$$

where  $m$  and  $n$  have various interpretations which depend on the null-hypothesis [6] and

$$(3.2) \quad c(s, m, n) = \pi^{1/2} \prod_{i=1}^s \frac{\Gamma\{\frac{1}{2}(2m + 2n + s + i + 2)\}}{\Gamma\{\frac{1}{2}(2m + i + 1)\} \Gamma\{\frac{1}{2}(2n + i + 1)\} \Gamma(\frac{1}{2}i)}.$$

Denote the sum of the roots  $\sum_i \theta_i$  by  $V_1^{(s)}$ , the subscript 1 indicating the first e.s.f. (Pillai uses the  $V^{(s)}$  notation.) Then the mathematical expectation of  $\exp(tV_1^{(s)})$  can easily be shown to take the determinantal form

$$(3.3) \quad E(e^{tV_1^{(s)}}) = \int \dots \int f(\theta_1, \dots, \theta_s) e^{tV_1^{(s)}} \prod_i d\theta_i$$

$$= c(s, m, n) \int_0^1 d\theta_s \int_0^{\theta_s} d\theta_{s-1} \dots \int_0^{\theta_2} d\theta_1$$

$$\cdot |\theta_{s-i+1}^{m+s-j} (1 - \theta_{s-i+1})^n e^{t\theta_{s-i+1}}|, \quad i, j = 1, \dots, s,$$

where the  $(i, j)$ th element of the determinantal expression is given by  $\theta_{s-i+1}^{m+s-j} (1 - \theta_{s-i+1})^n e^{t\theta_{s-i+1}}$ , by first noting that the product  $\prod_{i>j} (\theta_i - \theta_j)$  of  $f(\theta_1, \dots, \theta_s)$  is a Vandermonde determinant like (2.3) and then multiplying the  $i$ th row of this determinant by  $\theta_{s-i+1}^m (1 - \theta_{s-i+1})^n e^{t\theta_{s-i+1}}$ ,  $i = 1, \dots, s$ .

Now, denote (3.3) by  $U(m + s - 1, \dots, m; n; t)$  and more conveniently by  $U(s - 1, s - 2, \dots, 0)$  with the  $m$ 's and  $n$  omitted when  $t = 0$ . To obtain in general the moments of the sum of the roots in determinantal form,  $U(m + s - 1, \dots, m; n; t)$  is differentiated [5, 8] successively with respect to  $t$  and  $t$  is set equal to zero after each differentiation. The lower-order moments through the fourth moment are given by equations (3.2) through (3.5) of [8]. For the purpose of this paper, the third moment is reproduced below with  $V^{(s)}$  changed to  $V_1^{(s)}$ ,

$$(3.4) \quad E[(V_1^{(s)})^3] = c(s, m, n)[U(s + 2, s - 2, s - 3, \dots, 1, 0)$$

$$+ 2U(s + 1, s - 1, s - 3, \dots, 1, 0)$$

$$+ U(s, s - 1, s - 2, s - 4, \dots, 1, 0)].$$

**4. An alternative derivation of the moments of  $V_1^{(s)}$ .** We indicate in this section an alternative way to derive the moments of  $V_1^{(s)}$ , and we extend this method



later in Section 6 to obtain the moments of  $V_j^{(s)}$ , the  $j$ th e. s. f. of  $s$  roots, which are not yet available in the current literature except for  $j = 1$  and  $s$ .

Consider the classes of functions of  $\theta_1, \dots, \theta_s$  of form

$$(4.1) \quad U'(q_s, q_{s-1}, \dots, q_1; t) = |\theta_{s-i+1}^{q_s-j+1} e^{t\theta_{s-i+1}}|, \quad q_s > \dots > q_1 \geq m,$$

where  $i, j = 1, \dots, s$ , and are respectively the indices of the rows and columns of the determinantal expression. Denote those of form (4.1) for  $t = 0$  and  $q'_i = q_i - m$  simply by  $U'(q'_s, q'_{s-1}, \dots, q'_1)$ . Then  $U(m+s-1, m+s-2, \dots, m; n; t)$  of Section 3, i.e., (3.3), may be rewritten as

$$(4.2) \quad c(s, m, n) \int_0^1 d\theta_s \int_0^{\theta_s} d\theta_{s-1} \int_0^{\theta_{s-1}} \dots \int_0^{\theta_2} d\theta_1 \\ \cdot U'(m+s-1, \dots, m; t) \cdot \prod_{i=1}^s (1 - \theta_i)^n.$$

The class  $U(q_s, q_{s-1}, \dots, q_1; n; t)$  of functions of  $\theta_1, \dots, \theta_s$  generated by the successive differentiations of  $U(m+s-1, m+s-2, \dots, m; n; t)$  with respect to  $t$  can be represented by the class  $U'(q_s, q_{s-1}, \dots, q_1; t)$  of functions generated by successive differentiations of  $U'(m+s-1, \dots, m; t)$ .

Since  $U'(m+s-1, \dots, m; t)$  may be verified to be equal to

$$\prod_{i=1}^s \theta_i^m e^{tV_1^{(s)}} \prod_{i>j} (\theta_i - \theta_j)$$

by comparing (3.3) and (4.2),

$$(4.3) \quad E[(V_1^{(s)})^r] = d^r/dt^r \{E(e^{tV_1^{(s)}})\}|_{t=0} \\ = c(s, m, n) \int \dots \int \left\{ \prod_{i=1}^s \theta_i^m (V_1^{(s)})^r \prod_{i>j} (\theta_i - \theta_j) \right\} \left\{ \prod_i (1 - \theta_i)^n d\theta_i \right\},$$

where the right side of the first equality indicates evaluation of the  $r$ th derivative at  $t = 0$ . Obviously, the factor  $(V_1^{(s)})^r \prod_{i>j} (\theta_i - \theta_j)$  is a linear combination of functions in the class  $U'(q'_s, q'_{s-1}, \dots, q'_1)$  and so moments of  $V_1^{(s)}$  may be derived alternatively by finding the necessary linear combination in this class with the aid of the theorem in Section 2 applied in the reverse manner.

To illustrate now the alternative method of obtaining the moments of  $V_1^{(s)}$ , take the case of the third moment given by (3.4). Let  $\Phi_p$  be the equivalent c. h. s. f. of  $\phi_p$  in Section 2 when the arguments in  $x$ 's are replaced by arguments in  $\theta$ 's. The initial choice of the  $s$ -order determinant in the class  $U'(q'_s, q'_{s-1}, \dots, q'_1)$  is suggested by  $\Phi_1^3$  which is equivalent to  $(V_1^{(s)})^3$ . Choose the  $U'$ -determinant such that we have the elements  $\Phi_1, \Phi_1, \Phi_1, \Phi_0, \dots, \Phi_0$  along the principal diagonal. The product of these diagonal elements is  $\Phi_1^3$  since  $\Phi_0 = 1$  by definition. We have

$$(4.4) \quad \begin{vmatrix} \Phi_1 & \Phi_0 & 0 & 0 & \cdots & 0 \\ \Phi_2 & \Phi_1 & \Phi_0 & 0 & \cdots & 0 \\ \Phi_3 & \Phi_2 & \Phi_1 & 0 & \cdots & 0 \\ \Phi_4 & \Phi_3 & \Phi_2 & \Phi_0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Phi_s & \Phi_{s-1} & \Phi_{s-2} & \Phi_{s-4} & \cdots & \Phi_0 \end{vmatrix} = \Phi_1^3 + \Phi_3 - 2\Phi_1\Phi_2.$$

It may be remarked that if the determinant of (4.4) is multiplied by  $\prod_{i>j}(\theta_i - \theta_j)$  and the theorem of Section 2 is applied, the determinant reduces to  $U'(s, s-1, s-2, s-4, \cdots, 1, 0)$ .

We next wish to eliminate the product  $\Phi_1\Phi_2$  in the right-hand side of (4.4). This suggests taking the  $s$ -order determinant in the  $U'$ -class with elements in the principal diagonal given by  $\Phi_2, \Phi_1, \Phi_0, \cdots, \Phi_0$ . Thus

$$(4.5) \quad \begin{vmatrix} \Phi_2 & \Phi_0 & 0 & \cdots & 0 \\ \Phi_3 & \Phi_1 & 0 & \cdots & 0 \\ \Phi_4 & \Phi_2 & \Phi_0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Phi_{s+1} & \Phi_{s-1} & \Phi_{s-3} & \cdots & \Phi_0 \end{vmatrix} = \Phi_1\Phi_2 - \Phi_3.$$

By the same principle as (4.4) above, the left determinant of (4.5) can be reduced to  $U'(s+1, s-1, s-3, \cdots, 1, 0)$ . Finally, after inspecting right sides of (4.4) and (4.5), we need

$$(4.6) \quad \begin{vmatrix} \Phi_3 & 0 & 0 & \cdots & 0 \\ \Phi_4 & \Phi_0 & 0 & \cdots & 0 \\ \Phi_5 & \Phi_1 & \Phi_0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Phi_{s+2} & \Phi_{s-2} & \Phi_{s-3} & \cdots & \Phi_0 \end{vmatrix} = \Phi_3,$$

which gives  $U'(s+2, s-2, s-3, \cdots, 1, 0)$ . Combining properly the right-hand sides of (4.4) through (4.6), we see the equivalence

$$(4.7) \quad \Phi_1^3 = (\Phi_1^3 + \Phi_3 - 2\Phi_1\Phi_2) + 2(\Phi_1\Phi_2 - \Phi_3) + \Phi_3.$$

Now multiply (4.4) through (4.6) by  $\prod_{i=1}^s \theta_i^m (1 - \theta_i)^n \prod_{i>j} (\theta_i - \theta_j)$  and integrate over proper limits. We have

$$(4.8) \quad \begin{aligned} E(\Phi_1)^3 &= c(s, m, n) \int \cdots \int \{ U'(s, s-1, s-2, s-4, \cdots, 1, 0) \\ &+ 2U'(s+1, s-1, s-3, \cdots, 1, 0) \\ &+ U'(s+2, s-2, s-3, \cdots, 1, 0) \} \left\{ \prod_i \theta_i^m (1 - \theta_i)^n d\theta_i \right\}, \end{aligned}$$

which reduces to (3.4) after noting how the  $U$ -class there and the  $U'$ -class here have been defined. It may be remarked that the linear combination in the  $U'$ -class is really  $(V_1^{(s)})^3 \prod_{i>j} (\theta_i - \theta_j)$  by comparison with (4.3).

**5. The  $k$ th compound of a matrix.** In order to extend our results to the moments of  $V_j^{(s)}$ ,  $j = 2, \dots, s$ , we need an important property of compound matrices.

Consider the expansion of  $[\prod_{i=1}^k (1 - x_i t)]^{-1}$  into a power series

$$(5.1) \quad \left[ \prod_i (1 - x_i t) \right]^{-1} = 1 + \phi_1 t + \phi_2 t^2 + \dots + \phi_k t^k + \dots$$

where  $\phi_i$  is a c.h.s.f. in  $k$  arguments. Let  $a_j = \sum x_1 \dots x_j$  be the  $j$ th e.s.f. with  $k$  arguments in  $x$ 's. Multiplying both sides of (5.1) by  $\prod_i (1 - x_i t) = \sum_{j=0}^s (-1)^j a_j t^j$  and equating coefficients, we have

$$(5.2) \quad \begin{aligned} \phi_1 - a_1 &= 0 \\ \phi_2 - \phi_1 a_1 + a_2 &= 0 \\ \phi_3 - \phi_2 a_1 + \phi_1 a_2 - a_3 &= 0 \\ &\dots \dots \dots \\ \phi_k - \phi_{k-1} a_1 + \dots + (-1)^k a_k &= 0. \end{aligned}$$

If we define two  $(k+1)$ th order triangular matrices by

$$(a) = \begin{pmatrix} a_0 & 0 & \dots & 0 \\ -a_1 & a_0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ (-1)^k a_k & (-1)^{k-1} a_{k-1} & \dots & a_0 \end{pmatrix}$$

$$(\phi) = \begin{pmatrix} \phi_0 & 0 & \dots & 0 \\ \phi_1 & \phi_0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \phi_k & \phi_{k-1} & \dots & \phi_0 \end{pmatrix}$$

where  $a_0 = \phi_0 = 1$ , then it may be checked that an alternative form of (5.2) is

$$(5.3) \quad (a)(\phi) = (\phi)(a) = \mathbf{I}$$

where  $\mathbf{I}$  is a unit matrix of order  $(k+1)$ .

Now, consider the  $k$ th compound of a matrix  $(b)$ , denoted by  $(b)^{(k)}$ , defined by a matrix whose elements are  $k$ -order minors of  $\det(b)$  arranged in Aitken's lexical sense (see [1], p. 90), i.e., minors which come from the same group of  $k$  rows from  $(b)$  are placed in the same rows in  $(b)^{(k)}$ , the order being decided by the columns of  $(b)$  that the minors contain in the same manner that words are arranged in a dictionary. For instance, the minor containing columns 1, 3, 4 of  $(b)$  is preceded in the row of  $(b)^{(3)}$  by those minors containing columns 1, 2,  $q$  for  $q \geq 3$ . The same definition holds if the words *row, rows* are replaced by words *column, columns*, respectively, and *vice versa*.

Further, define the  $k$ th adjugate compound of  $(b)$ , denoted by  $\text{adj}^{(k)}(b)$ , as the transpose of the matrix formed from  $(b)^{(k)}$  after replacing every element in  $(b)^{(k)}$  by its cofactor in  $|b|$ .

It may be noted that from the way  $(b)^{(k)}$  and  $\text{adj}^{(k)}(b)$  are defined above,

$(b)^{(1)} = (b)$  and  $\text{adj}^{(1)}(b) = \text{adj}(b)$ . Furthermore, every element in the product  $(b)^{(k)} \text{adj}^{(k)}(b)$  is a Laplace Development according to  $k$ -ordered minors and their cofactors in  $\det(b)$ . It may be checked easily that only the diagonal elements in the product  $(b)^{(k)} \text{adj}^{(k)}(b)$  are expansions in minors by their *algebraic complements* (for definition, see [2], p. 23) and each is equal to  $\det(b)$ . The off-diagonal elements are sums of products of minors by the algebraic complements of some other minors and each sum is therefore equal to zero. Hence, if  $(b)$  is of order  $n$ ,

$$(5.4) \quad (b)^{(k)} \text{adj}^{(k)}(b) = |b| \mathbf{I},$$

where  $\mathbf{I}$  is a unit matrix of order  $n!/k!(n-k)!$ .

Consider now the product  $(a)(\phi)$  in (5.3). On applying the Binet-Cauchy theorem (see [1], p. 93) and multiplying both sides of the equation by  $(\text{adj}(a))^{(k)}$ , we have

$$(5.5) \quad (\phi)^{(k)}(a)^{(k)}(\text{adj}(a))^{(k)} = (\text{adj}(a))^{(k)} \mathbf{I}.$$

It may be recalled that  $(a) \text{adj}(a) = |a| \mathbf{I}$  and by the Binet-Cauchy theorem, the  $k$ th compound is

$$(5.6) \quad (a)^{(k)}(\text{adj}(a))^{(k)} = |a|^k \mathbf{I}.$$

Furthermore, comparing (*op. cit.*, p. 98) the equality in (5.6) with (5.4) after replacing  $(b)$  by  $(a)$ , we have

$$(5.7) \quad (\text{adj}(a))^{(k)} = |a|^{k-1} \text{adj}^{(k)}(a).$$

Using (5.6) and (5.7) and noting that  $|a| = 1$ , (5.5) reduces finally to

$$(5.8) \quad \begin{aligned} (\phi)^{(k)} &= (\text{adj}(a))^{(k)} \\ &= \text{adj}^{(k)}(a). \end{aligned}$$

From the nature of the construction of  $(\phi)^{(k)}$  and  $\text{adj}^{(k)}(a)$ , the last equality of (5.8) reveals an inner relationship of minors of  $|\phi|$  and  $|a|$  which plays a key role in the next section.

If the columns of elements of  $(\phi)^{(k')}$ , which are  $k'$ -order minors of  $(k+1)$ -order determinant  $|\phi|$ , are labelled by their highest suffixes occurring in the columns and if the same method of labelling is used for the elements of  $\text{adj}^{(k')}(a)$  which are  $(k+1-k')$ -order minors of  $|a|$ , then the two sets of suffixes form a bicomplementary set with respect to the highest index. Specifically, we restrict our use of (5.8) only to those minors with consecutive suffixes in the columns.

For example, if an element of  $(\phi)^{(3)}$  has column indices labelled 4, 2, 1 then the indices missing in the sequence of numbers 4, 3, 2, 1, 0 are 3, 0. Thus the complementary indices with respect to the highest index 4 in the corresponding  $\text{adj}^{(3)}(a)$  have labels 1, 4 in reverse order. Hence,

$$\begin{vmatrix} \phi_2 & \phi_0 & 0 \\ \phi_3 & \phi_1 & \phi_0 \\ \phi_4 & \phi_2 & \phi_1 \end{vmatrix} = \begin{vmatrix} -a_3 & a_0 \\ a_4 & -a_1 \end{vmatrix}.$$

**6. The mathematical expectations of the e.s.f.'s.** From the property (5.8) and the theorem in Section 2, the inverse derivation of the moments of the first e.s.f. of  $s$  roots may now be extended to any e.s.f. As an illustration, take the second moment of the second e.s.f. for the case of  $s = 3$ .

If the arguments in  $x$ 's of the  $\phi$ 's and  $a$ 's of Section 5 are now replaced by arguments in  $\theta$ 's, then  $\phi_i \rightarrow \Phi_i$  and  $a_i \rightarrow V_i^{(3)} = V_i$ , say, with superscripts omitted if the meaning is clear from the context. The second moment suggests taking the  $V$ -determinant with  $V_2, V_2$  in the principal diagonal so that the corresponding  $\Phi$ -determinant is of order 3 (since  $s = 3$ ). Thus

$$(6.1) \quad \begin{vmatrix} V_2 & -V_1 \\ -V_3 & V_2 \end{vmatrix} = V_2^2 - V_1 V_3$$

which suggests to add to (6.1) a  $V$ -determinant with  $V_1, V_3$  in the principal diagonal. Thus, we have only

$$(6.2) \quad \begin{vmatrix} V_2 & -V_1 \\ -V_3 & V_2 \end{vmatrix} + \begin{vmatrix} -V_3 & V_2 \\ V_4 & -V_1 \end{vmatrix} = V_2^2$$

since  $V_4 = 0$  in the case of  $s = 3$ . Using (5.8) the equivalent  $\Phi$ -determinants of the  $V$ -determinants are

$$(6.3) \quad \begin{vmatrix} \Phi_2 & \Phi_1 & 0 \\ \Phi_3 & \Phi_2 & 0 \\ \Phi_4 & \Phi_3 & \Phi_0 \end{vmatrix} + \begin{vmatrix} \Phi_1 & \Phi_0 & 0 \\ \Phi_2 & \Phi_1 & \Phi_0 \\ \Phi_4 & \Phi_2 & \Phi_1 \end{vmatrix}$$

on noting that the subscripts (in reverse order) of the last rows of the  $V$ -determinants form a bicomplementary set with respect to the highest index 4 with the *missing* subscripts in the last rows of the  $\Phi$ -determinants.

The next obvious step to find  $E[(V_2^{(3)})^2]$  is to employ directly the method of Section 4 on (6.3). This gives

$$\begin{aligned} E[(V_2^{(3)})^2] &= c(3, m, n) \int \cdots \int \{U'(4, 3, 0) + U'(4, 2, 1)\} \left\{ \prod_i \theta_i^m (1 - \theta_i)^n d\theta_i \right\} \\ &= c(3, m, n)[U(4, 3, 0) + U(4, 2, 1)]. \end{aligned}$$

The author is presently tabulating the lower-order moments of  $V_i^{(s)}$ ,  $i = 2, 3, 4$  and  $s = 2, 3, 4$  for values of  $2m = -1(1)10(10)60(20)120$  and  $2n = 10(10)200$ . The results will be reported at some future time.

**7. Acknowledgment.** I am indebted to the referee for pointing out an error in the proof of the theorem (Section 2) in an earlier draft and for his other useful criticisms and suggestions. I am also indebted to Professor William Kruskal for his helpful comments while this paper was in preparation.

## REFERENCES

- [1] A. C. AITKEN, *Determinants and Matrices*, 9th ed., Oliver and Boyd, London, 1956.
- [2] M. BOCHER, *Introduction to Higher Algebra*, Macmillan Co., New York, 1907.
- [3] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238-249.
- [4] P. L. HSU, "On the distribution of the roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250-258.
- [5] T. A. MIJARES, "On the distribution of the sum of six roots of a matrix in multivariate analysis," unpublished M.S. thesis submitted in March, 1958, to the University of the Philippines Graduate School.
- [6] K. C. S. PILLAI, "Some new test criteria in multivariate analysis," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 117-121.
- [7] K. C. S. PILLAI, "Some results useful in multivariate analysis," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1106-1113.
- [8] K. C. S. PILLAI AND TITO A. MIJARES, "On the moments of the trace of a matrix and approximations to its distribution," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1135-1140.
- [9] S. N. ROY, "*p*-statistics or some generalizations in analysis of variance appropriate to multivariate problems," *Sankyā*, Vol. 4 (1939), pp. 381-396.



## VARIANCE COMPONENTS IN THE UNBALANCED 2-WAY NESTED CLASSIFICATION

BY S. R. SEARLE

*New Zealand Dairy Board, Wellington, New Zealand*

**Introduction.** Sampling variances of estimates of components of variance obtained from data that are balanced (having the same number of observations in all subclasses) are easily derived because the mean squares in the analysis of variance are independent and distributed as  $\chi^2$ . The variance component estimates are linear functions of the mean squares and their variances can be derived accordingly, although their distributions are, in general, unknown. When the data are not balanced, however, and there are unequal numbers of observations in the subclasses the mean squares are no longer independent and they do not have  $\chi^2$ -distributions. Methods of deriving expressions for the sampling variances of the variance component estimates are developed for these situations in an earlier paper [3] and applied to the 1-way classification. A second paper [4] gives these expressions for the 2-way factorial classification, and extension to the 2-way hierarchical (nested) classification is presented here.

**Model and analysis of variance.** The earlier work discussed sampling variances of variance component estimates obtained by Henderson's Method 1 [2] from data having unequal subclass numbers, based on the completely random model, namely Eisenhart's Model II, [1]. The same situation is considered here for the 2-way nested classification.

The linear model for an observation  $x_{ijk}$  is taken as

$$x_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk}$$

where  $\mu$  is the general mean,  $\alpha_i$  is the effect due to the  $i$ th main classification,  $\beta_{ij}$  is the effect due to the  $j$ th sub-class within the  $i$ th main classification, and  $e_{ijk}$  is the residual error term peculiar to  $x_{ijk}$ . We suppose the number of classes in the main classification is  $a$ , so that  $i = 1, \dots, a$ ; and that there are  $c_i$  sub-classes within each of these so that  $j = 1, \dots, c_i$ . The total number of such sub-classes will be represented by  $b$ , giving  $b = \sum_{i=1}^a c_i$ . The number of observations in the  $j$ th subclass of the  $i$ th class is taken as  $n_{ij}$ . All terms of the model (except  $\mu$ ) are assumed to be normally distributed random variables with zero means and variances  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$  and  $\sigma_e^2$ . These are the variance components to be estimated, along with the sampling variances of their estimates.

---

Received December 4, 1960; revised June 30, 1961.

The customary analysis of variance can be written as

*Analysis of Variance*

Term	d.f.	Sums of Squares
Between main classes	$a - 1$	$T_a - T_f$
Between subclasses within main classes	$b - a$	$T_{ab} - T_a$
Within subclasses	$N - b$	$T_o - T_{ab}$
Total	$N - 1$	$T_o - T_f$

where, with the customary notation for totals and means, namely

$$x_{i..} = \sum_j \sum_k x_{ijk}, \quad n_{i.} = \sum_j n_{ij} \quad \text{and} \quad \bar{x}_{i..} = x_{i..}/n_{i.}$$

we have the uncorrected sums of squares

$$T_a = \sum_i n_{i.} \bar{x}_{i..}^2,$$

$$T_{ab} = \sum_i \sum_j n_{ij} \bar{x}_{ij.}^2,$$

$$T_o = \sum_i \sum_j \sum_k x_{ijk}^2$$

and

$$T_f = N \bar{x}^2, \dots,$$

$N$  being the total number of observations,  $N = \sum_i \sum_j n_{ij}$ .

The variance components can be estimated by equating each line of the above analysis of variance (except that for "total") to its expected value. Denoting the resulting estimates as  $\sigma_a^2$ ,  $\sigma_b^2$  and  $\sigma_e^2$ , the equations for obtaining them are, as given in Section 10.17 of [5]

$$\begin{aligned} T_a - T_f &= (N - k_1) \sigma_a^2 + (k_{12} - k_3) \sigma_b^2 + (a - 1) \sigma_e^2 \\ (1) \quad T_{ab} - T_a &= (N - k_{12}) \sigma_b^2 + (b - a) \sigma_e^2 \\ T_o - T_{ab} &= (N - b) \sigma_e^2. \end{aligned}$$

The  $k$ 's are functions of the  $n_{ij}$ 's, namely

$$k_1 = \sum_i n_{i.}^2 / N$$

$$k_3 = \sum_i \sum_j n_{ij}^2 / N$$

and

$$k_{12} = \sum_i (\sum_j n_{ij}^2) / n_{i.}.$$

The notation here follows that used previously in [4].

**Variances and covariances required.** The within sub-classes sum of squares,  $T_o - T_{ab}$ , has a  $\chi^2$ -distribution with  $(N - b)$  degrees of freedom and hence the variance of  $\sigma_e^2$  is

$$(2) \quad \text{var}(\sigma_e^2) = 2\sigma_e^4/(N - b).$$

Furthermore,  $T_o - T_{ab}$  is distributed independently of  $T_a$ ,  $T_{ab}$  and  $T_f$  so that the covariances of  $\sigma_a^2$  and  $\sigma_b^2$  with  $\sigma_e^2$  are obtained directly as

$$(3) \quad \text{cov}(\sigma_b^2, \sigma_e^2) = -(b - a) \text{var}(\sigma_e^2)/(N - k_{12})$$

and

$$(4) \quad \begin{aligned} \text{cov}(\sigma_a^2, \sigma_e^2) &= -[(k_{12} - k_3) \text{cov}(\sigma_b^2, \sigma_e^2) + (a - 1) \text{var}(\sigma_e^2)]/(N - k_1) \\ &= [(k_{12} - k_3)(b - a)/(N - k_{12}) - (a - 1)] \text{var}(\sigma_e^2)/(N - k_1). \end{aligned}$$

This independence property is also used for obtaining the variances of  $\sigma_a^2$  and  $\sigma_b^2$  and the covariance between them as linear functions of  $\text{var}(\sigma_e^2)$  and the variances and covariances of  $T_a$ ,  $T_{ab}$  and  $T_f$ . Thus

$$(5) \quad \text{var}(\sigma_b^2) = \frac{\text{var}(T_{ab} - T_a) + (b - a)^2 \text{var}(\sigma_e^2)}{(N - k_{12})^2}$$

and

$$(6) \quad \begin{aligned} (N - k_1)^2(N - k_{12})^2 \text{var}(\sigma_a^2) &= \text{var}[(N - k_3)T_a - (k_{12} - k_3)T_{ab} - (N - k_{12})T_f] \\ &\quad + [(N - k_3)a - (k_{12} - k_3)b - (N - k_{12})]^2 \text{var}(\sigma_e^2) \end{aligned}$$

and

$$(7) \quad \begin{aligned} (N - k_1)(N - k_{12}) \text{cov}(\sigma_a^2, \sigma_b^2) &= \\ \text{cov}(T_a - T_f)(T_{ab} - T_a) &+ (a - 1)(b - a) \text{var}(\sigma_e^2) \\ &\quad - (N - k_{12})(k_{12} - k_3) \text{var}(\sigma_b^2). \end{aligned}$$

The second term in each of these expressions can be obtained from equation (2) and the first can be found as a linear function of the variances and covariances of  $T_a$ ,  $T_{ab}$  and  $T_f$ . These we now proceed to find.

**Matrix methods.** The sampling variance of a quadratic function,  $\mathbf{x}'F\mathbf{x}$ , of normally-distributed random variables represented by the vector  $\mathbf{x}$  is  $2\text{tr}(VF)^2$  where  $V$  is the variance-covariance matrix appropriate to the variables in  $\mathbf{x}$ . The covariance between two quadratics  $\mathbf{x}'F\mathbf{x}$  and  $\mathbf{x}'G\mathbf{x}$  is  $2\text{tr}(VFVG)$ . These results can be applied to obtain the terms needed for equations (5) through (7) using matrices similar to those employed in Searle, [4]. First we define square matrices  $U_{ij}$ ,  $U_i$ , and  $U_N$  of order  $n_{ij}$ ,  $n_i$ , and  $N$  respectively, with all elements equal to one. Square matrices of order  $N$  with  $U$ -matrices in the diagonal and zeros elsewhere are defined as  $D$ -matrices; thus  $D_{ab}$  has the matrices  $U_{ij}$  in its

diagonal, for all values of  $i$  and  $j$ , and  $D_a$  has the matrices  $U_i$  in its diagonal, for all values of  $i$ . The variance-covariance matrix of the  $N$  observations arrayed in order  $k = 1 \dots n_{ij}$  within  $j$ -classes within each  $i$ -class can now be expressed as

$$(8) \quad V = \sigma_a^2 D_a + \sigma_{\beta}^2 D_{ab} + \sigma_e^2 I,$$

$I$  being an identity matrix.

Defining  $C_{ab}$  and  $C_a$  similar to  $D_{ab}$  and  $D_a$  only with matrices  $(n_{ij})^{-1}U_{ij}$  and  $(n_i)^{-1}U_i$  in the diagonal enables the quadratics in the analysis of variance to be written as

$$T_a = \mathbf{x}' C_a \mathbf{x}$$

$$T_{ab} = \mathbf{x}' C_{ab} \mathbf{x}$$

and

$$T_f = \mathbf{x}' U_N \mathbf{x}.$$

Thus

$$(9) \quad \begin{aligned} \text{var}(T_a) &= 2 \text{tr}(V C_a)^2 \\ &= 2 \text{tr}(\sigma_a^2 D_a + \sigma_{\beta}^2 D_{ab} C_a + \sigma_e^2 C_a)^2 \end{aligned}$$

after substitution from (8). This is a quadratic in the variance components which can be expanded, through the special form of the matrices, in terms of the  $n_{ij}$ 's using the expressions

$$\begin{aligned} k_4 &= \sum_i \sum_j n_{ij}^3 & k_6 &= \sum_i (\sum_j n_{ij}^3)/n_i \\ k_6 &= \sum_i (\sum_j n_{ij}^2)^2/n_i & k_7 &= \sum_i (\sum_j n_{ij}^2)^2/n_i \\ k_8 &= \sum_i n_i (\sum_j n_{ij}^2) & k_9 &= \sum_i n_i^3 \end{aligned}$$

Thus

$$\text{var}(T_a) = 2(Nk_1\sigma_a^4 + k_7\sigma_{\beta}^4 + a\sigma_e^4 + 2Nk_3\sigma_a^2\sigma_{\beta}^2 + 2N\sigma_a^2\sigma_e^2 + 2k_{12}\sigma_{\beta}^2\sigma_e^2).$$

A similar procedure for the other terms in (5), (6) and (7) leads to the following results:

$$\begin{aligned} \text{var}(T_{ab}) &= 2 \text{tr}(V C_{ab})^2 \\ &= \text{var}(T_a) + 2[(Nk_3 - k_7)\sigma_{\beta}^4 + (b - a)\sigma_e^4 + 2(N - k_{12})\sigma_{\beta}^2\sigma_e^2] \\ \text{var}(T_f) &= 2 \text{tr}(V U_N)^2/N^2 \\ &= 2(k_1\sigma_a^2 + k_3\sigma_{\beta}^2 + \sigma_e^2)^2 \\ \text{cov}(T_a T_{ab}) &= 2 \text{tr}(V C_a V C_{ab}) \\ &= \text{var}(T_a) + 2(k_5 - k_7)\sigma_{\beta}^4 \end{aligned}$$

$$\begin{aligned}\text{cov}(T_a T_f) &= 2 \text{tr}(VC_a VU_N)/N \\ &= 2[(k_9/N)\sigma_a^4 + (k_8/N)\sigma_\beta^4 + \sigma_e^4 \\ &\quad + 2(k_8/N)\sigma_a^2 \sigma_\beta^2 + 2k_1 \sigma_a^2 \sigma_e^2 + 2k_3 \sigma_\beta^2 \sigma_e^2]\end{aligned}$$

$$\begin{aligned}\text{cov}(T_{ab} T_f) &= 2 \text{tr}(VC_{ab} VU_N)/N \\ &= \text{cov}(T_a T_f) + 2\sigma_\beta^4(k_4 - k_6)/N.\end{aligned}$$

**Results.** Substituting the above expressions into (5) leads, after simplification, to

$$\text{var}(\hat{\sigma}_a^2) = \frac{2(\lambda_1 \sigma_a^4 + \lambda_2 \sigma_\beta^4 + \lambda_3 \sigma_e^4 + 2\lambda_4 \sigma_a^2 \sigma_\beta^2 + 2\lambda_5 \sigma_a^2 \sigma_e^2 + 2\lambda_6 \sigma_\beta^2 \sigma_e^2)}{(N - k_1)^2(N - k_{12})^2}$$

where

$$\begin{aligned}\lambda_1 &= (N - k_{12})^2[k_1(N + k_1) - 2k_9/N], \\ \lambda_2 &= k_3[N(k_{12} - k_3)^2 + k_3(N - k_{12})^2] + (N - k_3)^2 k_7 \\ &\quad - 2(N - k_3)[(k_{12} - k_3)k_8 + (N - k_{12})k_6/N] \\ &\quad + 2(N - k_{12})(k_{12} - k_3)k_4/N, \\ \lambda_3 &= [(N - k_{12})^2(N - 1)(a - 1) - (N - k_3)^2(a - 1)(b - a) \\ &\quad + (k_{12} - k_3)^2(N - 1)(b - a)]/(N - b), \\ \lambda_4 &= (N - k_{12})^2[k_3(N + k_1) - 2k_8/N], \\ \lambda_5 &= (N - k_{12})^2(N - k_1),\end{aligned}$$

and

$$\lambda_6 = (N - k_{12})(N - k_3)(k_{12} - k_3).$$

Similarly, expression (6) becomes

$$\begin{aligned}\text{var}(\hat{\sigma}_\beta^2) &= \frac{2(k_7 + Nk_3 - 2k_6)\sigma_\beta^4 + 4(N - k_{12})\sigma_\beta^2 \sigma_e^2 + 2(b - a)(N - a)\sigma_e^4/(N - b)}{(N - k_{12})^2}\end{aligned}$$

and (7) reduces to

$$\begin{aligned}(N - k_1)(N - k_{12}) \text{cov}(\hat{\sigma}_a^2 \hat{\sigma}_\beta^2) &= 2[k_6 - k_7 + (k_8 - k_4)/N]\sigma_\beta^4 \\ &\quad + 2(a - 1)(b - a)\sigma_e^4/(N - b) - (N - k_{12})(k_{12} - k_3) \text{var}(\hat{\sigma}_\beta^2).\end{aligned}$$

These variances are in terms of the unknown variance components  $\sigma_a^2$ ,  $\sigma_\beta^2$  and  $\sigma_e^2$  so that estimation of the variances in any particular case is only possible by replacing the components in these formulae by their estimates.

**Balanced data.** The above formulae reduce to the well-known results for balanced data when all the  $n_{ij}$  are put equal, to  $n$ , say. Suppose that all levels of the main classification have  $c$  sub-classes so that  $b = ac$ . Then, for example,

$$\text{var}(\hat{\sigma}_\beta^2) = \frac{2(an^2 + acn^2 - 2an^3)\sigma_\beta^4 + 4an(c-1)\sigma_\beta^2\sigma_e^2 + 2a(c-1)\sigma_e^4/ac(n-1)}{a^2n^2(c-1)^2}$$

which reduces to

$$\text{var}(\hat{\sigma}_\beta^2) = \frac{2}{n^2} \left[ \frac{n\sigma_\beta^2 + \sigma_e^2}{a(c-1)} + \frac{\sigma_e^4}{ac(n-1)} \right].$$

This is the result obtained directly for the balanced case when  $T_{ab} - T_a$  and  $T_o - T_{ab}$  are distributed independently as  $\chi^2$  with  $a(c-1)$  and  $ac(n-1)$  degrees of freedom respectively. Their expectations, obtained from equation (1), are

$$E(T_{ab} - T_a) = a(c-1)(n\sigma_\beta^2 + \sigma_e^2)$$

and

$$E(T_o - T_{ab}) = ac(n-1)\sigma_e^2$$

and their variances equal twice the square of their expectations divided by their degrees of freedom. The variance of the estimate of  $\sigma_\beta^2$ , namely

$$\hat{\sigma}_\beta^2 = \frac{1}{n} \left[ \frac{T_{ab} - T_a}{a(c-1)} - \frac{T_o - T_{ab}}{ac(n-1)} \right],$$

is accordingly as shown above.

**Acknowledgment.** Sincere thanks are given to Dr. C. R. Henderson of Cornell University, Ithaca, N. Y., for support of this work from funds of a research grant entitled "Estimation of genetic parameters" from the National Science Foundation of the U. S. A.

#### REFERENCES

- [1] EISENHART, CHURCHILL, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1-21.
- [2] HENDERSON, C. R., "Estimation of variance and covariance components," *Biometrics*, Vol. 9 (1953), pp. 226-252.
- [3] SEARLE, S. R., "Matrix methods in variance and covariance components analysis," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 737-748.
- [4] SEARLE, S. R., "Sampling variances of estimates of components of variance," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 167-178.
- [5] SNEDECOR, G. W., *Statistical Methods*, 5th Ed., Iowa State College Press, Ames, Iowa, 1957.



## SOME MAIN-EFFECT PLANS AND ORTHOGONAL ARRAYS OF STRENGTH TWO<sup>1</sup>

BY SIDNEY ADDELMAN<sup>2</sup> AND OSCAR KEMPTHORNE

*Iowa State University*

**1. Summary.** In this paper we present a method of constructing main-effect plans for symmetrical factorial experiments which can accommodate up to  $[2(s^n - 1)/(s - 1) - 1]$  factors, each at  $s = p^m$  levels, where  $p$  is a prime, with  $2s^n$  treatment combinations. As main-effect plans are orthogonal arrays of strength two the method presented permits the construction of the orthogonal arrays  $(2s^n, 2[s^n - 1]/[s - 1] - 1, s, 2)$ .

**2. Introduction.** Let there be  $k$  factors each of which can assume  $s = p^m$  levels, where  $p$  is a prime number. An orthogonal array of strength  $d$ , of size  $N$ , with  $k$  constraints and  $s$  levels consists of a subset of  $N$  treatment combinations from an  $s^k$  factorial experiment with the property that all  $s^d$  treatment combinations corresponding to any  $d$  factors chosen from the  $k$  occur an equal number of times in the subset. The array may be denoted by  $(N, k, s, d)$ .

The concept of orthogonal arrays was first introduced by Rao [1]. He discussed the use of these arrays as fractionally replicated plans for symmetrical factorial experiments which permit the estimation of main-effects and interactions up to order  $(d - 2)$  when higher order interactions are negligible.

The plans for fractionally replicated symmetrical factorial experiments which are developed in this paper are orthogonal arrays of strength two. We call these plans main-effect plans because they permit orthogonal estimation of all the main-effects when the interactions are negligible.

The main-effect plans derivable from the system of confounding developed by Fisher [2] can be represented by the orthogonal arrays  $(s^n, (s^n - 1)/(s - 1), s, 2)$ . These plans fall within the class of optimum multifactorial designs which were considered by Plackett and Burman [3].

It has been shown by Bose [4] that the maximum number of factors that it is possible to accommodate in a symmetrical factorial experiment in which each factor occurs at  $s = p^m$  levels and each block is of size  $s^n$ , without confounding any  $d$ -factor or lower order interaction, is given by the maximum number of points that it is possible to choose in the finite projective geometry  $PG(n - 1, p^m)$  so that no  $d$  of the chosen points are conjoint. This is equivalent to showing that the maximum number of constraints  $k$  in the orthogonal array  $(s^n, k, s, d)$  is given by the maximum number of points it is possible to choose in

Received August 11, 1960; revised April 17, 1961.

<sup>1</sup> This research was supported in whole or in part by the United States Air Force under Contract No. AF 33(616)-5599, monitored by the Aeronautical Research Laboratory, Wright Air Development Division.

<sup>2</sup> Now at the Research Triangle Institute.

$\text{PG}(n-1, p^m)$  so that no  $d$  of the chosen points are conjoint. Clearly the maximum number of constraints in the orthogonal array  $(s^n, k, s, 2)$  is equal to the number of points of  $\text{PG}(n-1, p^m)$ . Thus the maximum value of  $k$  is  $(s^n - 1)/(s - 1)$ . These facts are relevant in view of the method of construction to be presented.

**3. Preliminary notation and lemmas.** The finite projective geometry  $\text{PG}(n-1, p^m)$  is a geometrical representation of  $n$  factors each at  $s = p^m$  levels and their generalized interactions. We shall represent these  $n$  factors by  $X_1, X_2, \dots, X_n$  and their generalized interactions by  $k_1X_1 + k_2X_2 + \dots + k_nX_n$  where the  $k_i$  can take on any value of the Galois field  $\text{GF}(p^m)$  and it is understood that the coefficient of the first factor appearing in an interaction is unity.

Let  $u_0, u_1, \dots, u_{s-1}$  represent the elements of  $\text{GF}(p^m)$  and let  $u_0^2, u_1^2, \dots, u_{s-1}^2$  represent the squares of the elements of  $\text{GF}(p^m)$ . We shall denote the set of squared elements of  $\text{GF}(p^m)$  by  $\text{GF}^2(p^m)$ . It is easily verified that apart from the 0 element the set  $\text{GF}^2(p^m)$  forms a cyclic Abelian group under multiplication. It follows from the cyclic property that (i) when  $p = 2$ ,  $\text{GF}^2(p^m)$  contains each of the elements of  $\text{GF}(p^m)$  and (ii) when  $p$  is an odd prime, the elements of  $\text{GF}^2(p^m)$  comprise a subset of  $\frac{1}{2}(s+1)$  distinct elements of  $\text{GF}(p^m)$ , where one element occurs once and  $\frac{1}{2}(s-1)$  elements are duplicated.

Consider one of the factors  $X_i$  in a main-effect plan in which each  $X_i$  has  $s$  levels, each occurring  $s^{n-1}$  times in a total of  $s^n$  treatment combinations. Let  $X_i^2$  be a pseudo-factor obtained by squaring the levels of  $X_i$ . We now present the following lemmas:

**LEMMA 1.** *When  $p$  is an odd prime,  $X_i^2 + kX_i$  ( $k$  an element of  $\text{GF}(p^m)$ ) contains  $\frac{1}{2}(s+1)$  distinct levels, one level occurring  $s^{n-1}$  times and  $\frac{1}{2}(s-1)$  levels occurring  $2s^{n-1}$  times in  $s^n$  treatment combinations.*

**LEMMA 2.** *When  $p = 2$ ,  $X_i^2$  contains each of the  $s$  levels  $s^{n-1}$  times.*

**LEMMA 3.** *When  $p = 2$ ,  $X_i^2 + kX_i$  ( $k$  any element of  $\text{GF}(p^m)$  except 0) contains  $\frac{1}{2}s$  distinct levels each occurring  $2s^{n-1}$  times.*

Lemma 3 can be proved as follows. Let  $x_i$  range over the elements of  $\text{GF}(p^m)$  which represent the  $s$  levels of  $X_i$ . As  $x_i$  ranges over the elements of the field so does  $x_i + k$  where  $k$  is an element of  $\text{GF}(p^m)$ . Also if  $x_i + k = x_j \pmod{2}$  then  $x_j + k = x_i \pmod{2}$ . Hence  $x_i(x_i + k) = x_i x_j$  and  $x_j(x_j + k) = x_i x_j$ . Thus whatever values of  $x_i(x_i + k)$  are achieved they are achieved for at least two values of  $x_i$ .

It will now be shown that the values of  $x_i(x_i + k)$  are achieved for exactly two values of  $x_i$ . Let  $y$  be the generator of the field and let  $x_i = y^\alpha$  and  $k = y^\beta$ . Thus  $x_i(x_i + k) = y^\alpha(y^\alpha + y^\beta)$ . Suppose that

$$y^\alpha(y^\alpha + y^\beta) = y^\gamma(y^\gamma + y^\beta)$$

where

$$y^\alpha \neq y^\gamma \quad \text{and} \quad y^\alpha + y^\beta \neq y^\gamma.$$

Hence

$$\begin{aligned}(y^\alpha)^2 + y^\alpha y^\beta &= (y^\gamma)^2 + y^\gamma y^\beta \\ (y^\alpha + y^\gamma)^2 + (y^\alpha + y^\gamma) y^\beta &= 0 \\ (y^\alpha + y^\gamma)(y^\alpha + y^\gamma + y^\beta) &= 0.\end{aligned}$$

This implies that either  $y^\alpha + y^\gamma = 0$  and therefore  $y^\alpha = y^\gamma$  which is a contradiction or that  $y^\alpha + y^\gamma + y^\beta = 0$  and therefore  $y^\alpha + y^\beta = y^\gamma$  which is a contradiction. Hence the values of  $x_i(x_i + k)$  are achieved for exactly two values of  $x_i$  and Lemma 3 is proved.

LEMMA 4. The factor represented by  $X_i^2 + k_i X_i + \sum_{j \neq i} k_j X_j$ , ( $k_i$  and  $k_j$  elements of GF( $p^m$ )) where at least one  $k_i \neq 0$ , contains each of the  $s$  levels  $s^{n-1}$  times.

LEMMA 5. The levels of  $X_i^2 + k_1 X_i + k_2 X_j$  which occur in a plan with the  $u_i$  level of  $a_1 X_i + a_2 X_j$ , where  $k_1, k_2, a_1$  and  $a_2$  are elements of GF( $p^m$ ) and  $a_2 \neq 0$  are given by the values of  $x_i^2 + k_1 x_i + k_2 x_j + c(a_1 x_i + a_2 x_j) - cu_i$  where  $k_2 + ca_2 = 0$  and  $x_i$  ranges over the elements of GF( $p^m$ ).

PROOF. When  $a_1 X_i + a_2 X_j$  takes on the  $u_i$  level then  $a_1 x_i + a_2 x_j = u_i$  and thus

$$x_j = (u_i - a_1 x_i)/a_2.$$

Hence the levels of the factor  $X_i^2 + k_1 X_i + k_2 X_j$  which occur with the level  $u_i$  of  $a_1 X_i + a_2 X_j$  can be represented by

$$\begin{aligned}x_i^2 + k_1 x_i + k_2 x_j &= x_i^2 + k_1 x_i + k_2(u_i - a_1 x_i)/a_2 \\ &= x_i^2 + (k_1 - k_2 a_1/a_2)x_i + (k_2/a_2)u_i.\end{aligned}$$

Since  $k_2 + ca_2 = 0$ , then  $c = -k_2/a_2$ . Thus

$$x_i^2 + (k_1 - k_2 a_1/a_2)x_i + (k_2/a_2)u_i = x_i^2 + k_1 x_i + k_2 x_j + c(a_1 x_i + a_2 x_j) - cu_i,$$

and the lemma is proved.

Two factors  $X_i$  and  $X_j$  are said to be orthogonal to each other if each level of  $X_j$  occurs the same number of times with every level of  $X_i$ . Two factors  $X_i$  and  $X_j$  are said to be semi-orthogonal to each other if (i) for  $p$  an odd prime, one level of  $X_j$  occurs  $s^{n-2}$  times and  $\frac{1}{2}(s-1)$  levels of  $X_j$  each occur  $2s^{n-2}$  times with each level of  $X_i$  and (ii) for  $p = 2$ ,  $\frac{1}{2}s$  levels of  $X_j$  each occur  $2s^{n-2}$  times with each level of  $X_i$ .

It follows from Lemmas 1, 3, and 5 that when  $p$  is an odd prime or when  $k_1 - k_2 a_1/a_2 \neq 0$ , then  $a_1 X_i + a_2 X_j$  is semi-orthogonal to  $X_i^2 + k_1 X_i + k_2 X_j$ . It follows from Lemmas 2 and 5 that when  $p = 2$  and  $k_1 - k_2 a_1/a_2 = 0$  then  $a_1 X_i + a_2 X_j$  is orthogonal to  $X_i^2 + k_1 X_i + k_2 X_j$ . Employing an argument similar to that used in Lemma 5 it can be deduced that  $kX_i^2 + k_1 X_i + X_j$  and  $kX_i^2 + k_2 X_i + X_j$  are orthogonal to each other when  $k_1 \neq k_2$ .

Lemma 5 can be generalized to include more than two factors as stated in Lemma 5a.

LEMMA 5a. The levels of  $X_i^2 + k_i X_i + \sum_{j \neq i} k_j X_j$  which occur in a plan with the  $u_i$  level of  $a_i X_i + \sum_{j \neq i} a_j X_j$  are given by the values of

$$x_i^2 + k_i x_i + \sum_{j \neq i} k_j x_j + c(a_i x_i + \sum_{j \neq i} a_j x_j) - cu_i$$

where  $k_j + ca_j = 0$  for all  $j \neq i$ . If the  $a_j$  and the  $k_j$  are not of such a form that  $k_j + ca_j = 0$  for all  $j \neq i$  and some  $c$  contained in  $\text{GF}(p^m)$  then the two factors are orthogonal.

LEMMA 6. When  $p$  is a prime the complements in  $\text{GF}(p^m)$  to the elements in  $\text{GF}^2(p^m)$  are the set of elements in  $\text{GF}^2(p^m)$  each multiplied by an element of  $\text{GF}(p^m)$  which is not an element of  $\text{GF}^2(p^m)$ . If the set of elements in  $\text{GF}^2(p^m)$  and their set of complements are taken together in one set, the elements of  $\text{GF}(p^m)$  are obtained.

PROOF. From abstract group theory (see Birkhoff and Mac Lane [5]) we employ a lemma which states that two right cosets of a subgroup are either identical or without common elements. Now the elements of  $\text{GF}^2(p^m)$  form an Abelian subgroup of the elements of  $\text{GF}(p^m)$ . Hence multiplying each element of  $\text{GF}^2(p^m)$  by an element of  $\text{GF}(p^m)$  which is not an element of  $\text{GF}^2(p^m)$  yields the complementary set to  $\text{GF}^2(p^m)$ .

It is clear from Lemma 2 that when  $p = 2$  the set complementary to  $\text{GF}^2(p^m)$  is the null set.

#### 4. Construction of main-effect plans.

THEOREM 1. There exists a main-effect plan for  $[2(s^n - 1)/(s - 1) - 1]$  factors, each at  $s = p^m$  levels, with  $2s^n$  treatment combinations.

PROOF. In order to facilitate the presentation of the proof of Theorem 1, let  $n = 2$ . First construct an orthogonal main effect plan for  $(s^2 - 1)/(s - 1)$  factors each at  $s$  levels in  $s^2$  trials, represented by the two factors  $X_1$  and  $X_2$  and their generalized interactions  $X_1 + X_2, X_1 + 2X_2, \dots, X_1 + (s - 1)X_2$ , where the coefficients  $1, 2, \dots, (s - 1)$  are elements of  $\text{GF}(p^m)$ , addition and multiplication being performed within this field. To these add

$$[(s^2 - 1)/(s - 1) - 1]$$

factors represented by

$$X_1^2 + X_2, X_1^2 + X_1 + X_2, X_1^2 + 2X_1 + X_2, \dots, X_1^2 + (s - 1)X_1 + X_2.$$

These  $[2(s^n - 1)/(s - 1) - 1]$  factors in  $s^n$  observations represent the first half of the main-effect plan.

Note from the preceding lemmas that when  $p$  is a prime number,  $X_1 + a_i X_2$  and  $X_1^2 + k_i X_1 + X_2$  are semi-orthogonal and also that  $X_2$  and  $X_1^2 + k_i X_1 + X_2$  are semi-orthogonal for all  $a_i$  and  $k_i$  in  $\text{GF}(p^m)$  except  $a_i = 0$ . All other pairs of factors are clearly orthogonal. If  $p = 2$  and  $(k_i - a_i/a_i) = 0$ , then  $a_i X_1 + a_i X_2$  and  $X_1^2 + k_i X_1 + X_2$  are orthogonal.

The second half of the plan is chosen so that the pairs of factors which are orthogonal in the first half are also orthogonal in the second half and pairs of

factors which are semi-orthogonal in the first half are semi-orthogonal in a complementary manner in the second half. The factors in the second half which correspond to the factors of the first half can be denoted by

$$X_1, X_2, X_1 + X_2 + b_1, X_1 + 2X_2 + b_2, \dots, X_1 + (s-1)X_2 + b_{s-1},$$

$$kX_1^2 + X_2, kX_1^2 + k_1X_1 + X_2 + c_1,$$

$$kX_1^2 + k_2X_1 + X_2 + c_2, \dots, kX_1^2 + k_{s-1}X_1 + X_2 + c_{s-1}$$

where the coefficients  $b_1, b_2, \dots, b_{s-1}, k, k_1, k_2, \dots, k_{s-1}, c_1, c_2, \dots, c_{s-1}$ , which are to be determined, are elements of  $\text{GF}(p^m)$ .

From Lemma 5, it is seen that the levels of  $X_1^2 + X_2$  which occur with the  $u_i$  level of  $X_2$  are given by the values of  $x_1^2 + u_i$  where  $x_1$  takes on the values of the elements of  $\text{GF}(p^m)$ . Without loss of generality we may let  $u_i = u_0 = 0$ . When  $p$  is an odd prime, the values of  $kX_1^2 + X_2$ , where  $k$  is an element of  $\text{GF}(p^m)$  but not an element of  $\text{GF}^2(p^m)$ , which occur with the  $u_i = 0$  level of  $X_2$  are given by the values of  $kx_1^2$ . As shown in Lemma 6,  $kx_1^2$  complements  $x_1^2$ .

Thus, when  $p$  is an odd prime  $k$  can take on the value of any element in  $\text{GF}(p^m)$  which is not an element of  $\text{GF}^2(p^m)$ . If  $p = 2$  it is clear from Lemma 2 that  $k = 1$ .

A method for determining the constants  $b_1, b_2, \dots, b_{s-1}, k_1, k_2, \dots, k_{s-1}, c_1, c_2, \dots, c_{s-1}$ , when  $s = p^m$  and  $p$  is an odd prime is now presented. In order that the levels of  $kX_1^2 + X_2$  which occur with the 0 level of  $X_1 + a_iX_2 + b_i$  be the complements of the levels of  $X_1^2 + X_2$  which occur with the 0 levels of  $X_1 + a_iX_2, b_i$  must be such that the values which  $kx_1^2 - (1/a_i)x_1 - b_i/a_i$  takes when  $x_1$  ranges over the field  $\text{GF}(p^m)$  complements the values which  $x_1^2 - (1/a_i)x_1$  takes. Now  $x_1^2 - (1/a_i)x_1$  consists of one element of  $\text{GF}(p^m)$  occurring once and  $\frac{1}{2}(s-1)$  elements occurring twice. Let the unique element of  $\text{GF}(p^m)$  be  $u_1$ . Then  $x_1^2 - (1/a_i)x_1 = u_1$  must have only one solution as  $x_1$  ranges over the elements of  $\text{GF}(p^m)$ . Thus  $1/a_i^2 + 4u_1 = 0$  and hence  $4u_1 = -1/a_i^2$ . Since  $kx_1^2 - (1/a_i)x_1 - b_i/a_i$  must complement  $x_1^2 - (1/a_i)x_1$ , the equation

$$kx_1^2 - (1/a_i)x_1 - b_i/a_i = u_1$$

must also have only one solution. Therefore

$$1/a_i^2 + 4k(b_i/a_i + u_1) = 0.$$

Substituting  $4u_1 = -1/a_i^2$  in this equation and solving for  $b_i$  we get

$$(1) \quad b_i = (k-1)/4ka_i.$$

To find the levels of  $X_1^2 + d_iX_1 + X_2$  which occur with the 0 levels of  $X_2$  note that there exists an element of  $\text{GF}(p^m)$ ,  $u_2$  say, such that  $x_1^2 + d_ix_1 = u_2$  has only one solution.

Thus  $d_i^2 + 4u_2 = 0$  and hence  $4u_2 = -d_i^2$ . In order that the levels of  $kX_1^2 + k_iX_1 + X_2 + c_i$  which occur with the 0 levels of  $X_2$  complement those given by

$x_1^2 + d_i x_1$ , then  $kx_1^2 + k_i x_1 + c_i = u_2$  must have only one solution. Substituting  $4u_2 = -d_i^2$  in this equation and solving for  $c_i$  we get

$$(2) \quad c_i = k_i^2/4k - d_i^2/4.$$

To find the levels of  $X_1^2 + d_i X_1 + X_2$  which occur with the 0 levels of  $X_1 + a_i X_2$  note that there exists an element of GF( $p^m$ ),  $u_3$  say, such that  $x_1^2 + (d_i - 1/a_i)x_1 = u_3$  has only one solution. Thus

$$(d_i - 1/a_i)^2 + 4u_3 = 0 \text{ and } 4u_3 = -(d_i - 1/a_i)^2.$$

Since  $kx_1^2 + (k_1 - 1/a_i)x_1 + (c_i - b_i/a_i)$  must complement  $x_1^2 + (d_i - 1/a_i)x_1$ , the equation

$$kx_1^2 + (k_1 - 1/a_i)x_1 + (c_i - b_i/a_i) = u_3$$

must also have only one solution as  $x_1$  ranges over the elements of GF( $p^m$ ). Therefore

$$(k_i - 1/a_i)^2 - 4k[(c_i - b_i/a_i) - u_3] = 0.$$

Substituting  $4u_3 = -(d_i - 1/a_i)^2$  and equations (1) and (2) into this equation we get

$$(3) \quad k_i = kd_i.$$

Hence equation (2) can be rewritten as

$$(4) \quad c_i = d_i^2/4(k - 1).$$

Thus  $k$  is determined by choosing an element of GF( $p^m$ ) which is not an element of GF<sup>2</sup>( $p^m$ ). By letting  $a_i = 1, 2, \dots, s-1$  we can determine  $b_1, b_2, \dots, b_{s-1}$  from equation (1). Then setting  $d_i = 1, 2, \dots, s-1$  we determine  $k_1, k_2, \dots, k_{s-1}$  from equation (3) and  $c_1, c_2, \dots, c_{s-1}$  from equation (4).

The procedure employed above cannot be applied when  $p = 2$  since  $x_1^2 + cx_1$  consists of  $\frac{1}{2}s$  elements of GF( $2^m$ ), each occurring twice. Thus there exists no element  $u$  such that  $x_1^2 + cx_1 = u$  must have only one solution.

We deduce from Lemma 2 that when  $p = 2$ , then  $k = 1$ . In order that the levels of  $X_1^2 + X_2$  which occur with the 0 levels of  $X_1 + a_i X_2 + b_i$  ( $a_i = 1, 2, 3, \dots, s-1$ ) complement the levels of  $X_1^2 + X_2$  which occur with the 0 levels of  $X_1 + a_i X_2$  then the levels given by  $x_1^2 - (1/a_i)x_1 - b_i/a_i$  must complement the levels given by  $x_1^2 - (1/a_i)x_1$  when  $x_1$  ranges over GF( $2^m$ ). It is easily verified that  $b_i$  can be any one of the  $2^{m-1}$  elements of GF( $2^m$ ) which are not given by  $x_1^2 - (1/a_i)x_1$ .

In order that the levels of  $X_1^2 + k_i X_1 + X_2 + c_i$  which occur with the 0 levels of  $X_2$  complement the levels of  $X_1^2 + d_i X_1 + X_2$  which occur with the 0 levels of  $X_2$ , then the values given by  $x_1^2 + k_i x_1 + c_i$  must complement the values given by  $x_1^2 + d_i x_1$ . It can be shown that  $k_i = d_i$  and  $c_i$  can be any one of the  $2^{m-1}$  elements of GF( $2^m$ ) which are not given by the values of  $x_1^2 + d_i x_1$ .



By finding the values of  $X_1^2 + k_i X_1 + X_2 + c_i$  which occur with the 0 levels of  $X_1 + a_i X_2 + b_i$  and which complement the values of  $X_1^2 + d_i X_1 + X_2$  that occur with the 0 levels of  $X_1 + a_i X_2$  a set of  $b_i$  and  $c_i$  which satisfy all the requirements to have the second half of the plan complement the first half of the plan can be determined.

When  $n > 2$  the same procedures will yield the desired plans if Lemma 5a is utilized in place of Lemma 5. Thus the theorem is proved.

**5. Examples.** Some of the more useful orthogonal arrays which can be constructed by the above procedures are: (18, 7, 3, 2), (54, 25, 3, 2), (32, 9, 4, 2), (128, 41, 4, 2), (50, 11, 5, 2), (250, 61, 5, 2), (98, 15, 7, 2), (128, 17, 8, 2) and (162, 19, 9, 2).

Bose and Bush [6] have constructed the arrays (18, 7, 3, 2) and (32, 9, 4, 2) by other procedures and have shown that  $[2(s^n - 1)/(s - 1) - 1]$  is the maximum number of constraints that arrays of size  $2s^n$  can accommodate.

We now present two examples of the construction of main effect plans for  $[2(s^n - 1)/(s - 1) - 1]$  factors each at  $s = p^m$  levels with  $2s^n$  treatment combinations. The first example illustrates the construction of a plan for eleven factors, each at five levels with fifty treatment combinations. This plan is the orthogonal array (50, 11, 5, 2).

The eleven factors which represent the first twenty-five treatment combinations are denoted by  $X_1, X_2, X_1 + X_2, X_1 + 2X_2, X_1 + 3X_2, X_1 + 4X_2, X_1^2 + X_2, X_1^2 + X_1 + X_2, X_1^2 + 2X_1 + X_2, X_1^2 + 3X_1 + X_2$  and  $X_1^2 + 4X_1 + X_2$ . The corresponding eleven factors representing the second half of the plan are denoted by  $X_1, X_2, X_1 + X_2 + b_1, X_1 + 2X_2 + b_2, X_1 + 3X_2 + b_3, X_1 + 4X_2 + b_4, kX_1^2 + X_2, kX_1^2 + k_1X_1 + X_2 + c_1, kX_1^2 + k_2X_1 + X_2 + c_2, kX_1^2 + k_3X_1 + X_2 + c_3$  and  $kX_1^2 + k_4X_1 + X_2 + c_4$ .

The elements of GF(5) are 0, 1, 2, 3 and 4. Hence the elements of  $GF^2(5)$  are 0, 1, 4, 4, 1. From Lemma 6, therefore,  $k = 2$  or  $k = 3$ . Let us choose  $k = 3$ . Hence, from equation (1)

$$b_i = 1/a_i.$$

Thus, when

$$a_i = 1 \quad \text{then} \quad b_1 = 1$$

$$a_i = 2 \quad \text{then} \quad b_2 = 3$$

$$a_i = 3 \quad \text{then} \quad b_3 = 2$$

$$a_i = 4 \quad \text{then} \quad b_4 = 4.$$

Now, from equations (3) and (4)

$$k_i = 3d_i \text{ and } c_i = 3d_i^2.$$

Thus, when

$$d_i = 1, \quad \text{then} \quad k_1 = 3, c_1 = 3$$

$$d_i = 2, \text{ then } k_2 = 1, c_2 = 2$$

$$d_i = 3, \text{ then } k_3 = 4, c_3 = 2$$

$$d_i = 4, \text{ then } k_4 = 2, c_4 = 3.$$

The eleven factor representations for the second half of the plan are therefore given by:  $X_1, X_2, X_1 + X_2 + 1, X_1 + 2X_2 + 3, X_1 + 3X_2 + 2, X_1 + 4X_2 + 4, 3X_1^2 + X_2, 3X_1^2 + 3X_1 + X_2 + 3, 3X_1^2 + X_1 + X_2 + 2, 3X_1^2 + 4X_1 + X_2 + 2$  and  $3X_1^2 + 2X_1 + X_2 + 3$ .

The second example will illustrate the construction of the plan for nine factors each at four levels with thirty-two treatment combinations. This plan is the orthogonal array (32, 9, 4, 2).

The nine factors which represent the first sixteen treatment combinations are denoted by  $X_1, X_2, X_1 + X_2, X_1 + 2X_2, X_1 + 3X_2, X_1^2 + X_2, X_1^2 + X_1 + X_2, X_1^2 + 2X_1 + X_2$  and  $X_1^2 + 3X_1 + X_2$ . The corresponding nine factors representing the second half of the plan are denoted by  $X_1, X_2, X_1 + X_2 + b_1, X_1 + 2X_2 + b_2, X_1 + 3X_2 + b_3, X_1^2 + X_2, X_1^2 + k_1X_1 + X_2 + c_1, X_1^2 + k_2X_1 + X_2 + c_2$  and  $X_1^2 + k_3X_1 + X_2 + c_3$ . The coefficients are elements of GF (2<sup>2</sup>) and all additions and multiplications are performed within this field.

Solving for  $c_1$  so that the levels of  $X_1^2 + X_2$  which occur with the 0 level of  $X_1 + X_2 + b_1$  complements the levels of  $X_1^2 + X_2$  which occur with the 0 level of  $X_1 + X_2$  we find that  $b_1 = 2$  or  $b_1 = 3$ . Similarly we find that  $b_2 = 1$  or  $b_2 = 2$  and that  $b_3 = 1$  or  $b_3 = 3$ .

As we wish the levels of  $X_1^2 + k_iX_1 + X_2 + c_i$  which occur with the 0 level of  $X_1 + a_iX_2$  to be complements to the levels of  $X_1^2 + k_iX_1 + X_2$  which occur with the 0 level of  $X_1 + a_iX_2$  we find that

$$k_1 = 1, \quad k_2 = 2, \quad k_3 = 3,$$

$$b_1 + c_2 = 1 \text{ or } 3, \quad b_1 + c_3 = 1 \text{ or } 2, \quad 3b_2 + c_1 = 1 \text{ or } 2,$$

$$3b_2 + c_2 = 2 \text{ or } 3, \quad 2b_3 + c_1 = 1 \text{ or } 3 \quad \text{and} \quad 2b_3 + c_3 = 2 \text{ or } 3.$$

Values of  $b_1, b_2, b_3, c_1, c_2$  and  $c_3$  which are consistent with all the above equations are  $b_1 = c_1 = 2, b_2 = c_2 = 1$  and  $b_3 = c_3 = 3$ . A second set of solutions is  $b_1 = c_1 = 3, b_2 = c_2 = 2$  and  $b_3 = c_3 = 1$ . These are the only two possible sets of solutions for this plan.

Since the coefficients satisfy all the properties required to make the second half of the plan complement the first half every pair of factors is orthogonal.

**6. Some useful orthogonal arrays.** In this final section we present the factors which represent the first and second halves of the arrays (18, 7, 3, 2) and (54, 25, 3, 2) and the treatment combinations which constitute the array (50, 11, 5, 2).

The factors representing the first half of the orthogonal array (18, 7, 3, 2) are:

$$X_1, X_2, X_1 + X_2, X_1 + 2X_2, X_1^2 + X_2, X_1^2 + X_1 + X_2, X_1^2 + 2X_1 + X_2.$$

The factors representing the second half of this array are:

$$X_1, X_2, X_1 + X_2 + 2, X_1 + 2X_2 + 1, 2X_1^2 + X_2, 2X_1^2 + 2X_1 + X_2 + 1, \\ 2X_1^2 + X_1 + X_2 + 1.$$

The factors representing the first half of the orthogonal array (54, 25, 3, 2) are:

$$X_1, X_2, X_1 + X_2, X_1 + 2X_2, X_3, X_1 + X_3, X_1 + 2X_3, X_2 + X_3, X_2 + 2X_3, \\ X_1 + X_2 + X_3, X_1 + X_2 + 2X_3, X_1 + 2X_2 + X_3, X_1 + 2X_2 + 2X_3, \\ X_1^2 + X_2, X_1^2 + X_1 + X_2, X_1^2 + 2X_1 + X_2, X_1^2 + X_3, X_1^2 + X_1 + X_3, \\ X_1^2 + 2X_1 + X_3, X_1^2 + X_2 + X_3, X_1^2 + X_1 + X_2 + X_3, X_1^2 + 2X_1 + X_2 + X_3, \\ X_1^2 + X_2 + 2X_3, X_1^2 + X_1 + X_2 + 2X_3, X_1^2 + 2X_1 + X_2 + 2X_3.$$

The factors representing the second half of this array are:

$$X_1, X_2, X_1 + X_2 + 2, X_1 + 2X_2 + 1, X_3, X_1 + X_3 + 2, X_1 + 2X_3 + 1, \\ X_2 + X_3, X_2 + 2X_3, X_1 + X_2 + X_3 + 2, X_1 + X_2 + 2X_3 + 2, \\ X_1 + 2X_2 + X_3 + 1, X_1 + 2X_2 + 2X_3 + 1, \\ 2X_1^2 + X_2, 2X_1^2 + 2X_1 + X_2 + 1, 2X_1^2 + X_1 + X_2 + 1, 2X_1^2 + X_3, \\ 2X_1^2 + 2X_1 + X_3 + 1, 2X_1^2 + X_1 + X_3 + 1, 2X_1^2 + X_2 + X_3, \\ 2X_1^2 + 2X_1 + X_2 + X_3 + 1, 2X_1^2 + X_1 + X_2 + X_3 + 1, 2X_1^2 + X_2 + 2X_3, \\ 2X_1^2 + 2X_1 + X_2 + 2X_3 + 1, 2X_1^2 + X_1 + X_2 + 2X_3 + 1.$$

The factors representing the orthogonal array (50, 11, 5, 2) were deduced in Section 5. The following fifty treatment combinations constitute a main-effect plan for eleven factors each at five levels and the array (50, 11, 5, 2). The treatment combinations are divided into two sets of twenty-five, the first set being the first half of the plan and the second set being the second half of the plan (see Table 1 following references).

#### REFERENCES

- [1] C. R. RAO, "Factorial experiments derivable from combinational arrangements of arrays," *J. Roy. Stat. Soc., Suppl.*, Vol. 9 (1947), pp. 128-139.
- [2] R. A. FISHER, "A system of confounding for factors with more than two alternatives giving completely orthogonal cubes and higher powers," *Ann. Eugenics*, Vol. 12 (1945), pp. 283-290.
- [3] R. L. PLACKETT AND J. P. BURMAN, "The design of optimum multifactorial experiments," *Biometrika*, Vol. 33 (1946), pp. 305-325.
- [4] R. C. BOSE, "Mathematical theory of symmetrical factorial designs," *Sankhyā*, Vol. 8 (1947), pp. 107-166.
- [5] G. BIRKHOFF AND S. MAC LANE, *A Survey of Modern Algebra*, Rev. ed., Macmillan Co., New York, 1953.
- [6] R. C. BOSE AND K. A. BUSH, "Orthogonal arrays of strength two and three," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 508-524.

TABLE 1

0	0	0	0	0	0	0	0	0	0	0
0	1	1	2	3	4	1	1	1	1	1
0	2	2	4	1	3	2	2	2	2	2
0	3	3	1	4	2	3	3	3	3	3
0	4	4	3	2	1	4	4	4	4	4
1	0	1	1	1	1	1	2	3	4	0
1	1	2	3	4	0	2	3	4	0	1
1	2	3	0	2	4	3	4	0	1	2
1	3	4	2	0	3	4	0	1	2	3
1	4	0	4	3	2	0	1	2	3	4
2	0	2	2	2	2	4	1	3	0	2
2	1	3	4	0	1	0	2	4	1	3
2	2	4	1	3	0	1	3	0	2	4
2	3	0	3	1	4	2	4	1	3	0
2	4	1	0	4	3	3	0	2	4	1
3	0	3	3	3	3	4	2	0	3	1
3	1	4	0	1	2	0	3	1	4	2
3	2	0	2	4	1	1	4	2	0	3
3	3	1	4	2	0	2	0	3	1	4
3	4	2	1	0	4	3	1	4	2	0
4	0	4	4	4	4	1	0	4	3	2
4	1	0	1	2	3	2	1	0	4	3
4	2	1	3	0	2	3	2	1	0	4
4	3	2	0	3	1	4	3	2	1	0
4	4	3	2	1	0	0	4	3	2	1
0	0	1	3	2	4	0	3	2	2	3
0	1	2	0	0	3	1	4	3	3	4
0	2	3	2	3	2	2	0	4	4	0
0	3	4	4	1	1	3	1	0	0	1
0	4	0	1	4	0	4	2	1	1	2
1	0	2	4	3	0	3	4	1	4	3
1	1	3	1	1	4	4	0	2	0	4
1	2	4	3	4	3	0	1	3	1	0
1	3	0	0	2	2	1	2	4	2	1
1	4	1	2	0	1	2	3	0	3	2
2	0	3	0	4	1	2	1	1	2	4
2	1	4	2	2	0	3	2	2	3	0
2	2	0	4	0	4	4	3	3	4	1
2	3	1	1	3	3	0	4	4	0	2
2	4	2	3	1	2	1	0	0	1	3
3	0	4	1	0	2	2	4	2	1	1
3	1	0	3	3	1	3	0	3	2	2
3	2	1	0	1	0	4	1	4	3	3
3	3	2	2	4	4	0	2	0	4	4
3	4	3	4	2	3	1	3	1	0	0
4	0	0	2	1	3	3	3	4	1	4
4	1	1	4	4	2	4	4	0	2	0
4	2	2	1	2	1	0	0	1	3	1
4	3	3	3	0	0	1	1	2	4	2
4	4	4	0	3	4	2	2	3	0	3

# ON A GEOMETRICAL METHOD OF CONSTRUCTION OF PARTIALLY BALANCED DESIGNS WITH TWO ASSOCIATE CLASSES

By ESTHER SEIDEN

*Michigan State University*

**1. Introduction.** The method of construction of partially balanced block designs discussed here is different than the ones known in the literature. It is based on the existence of an oval (maximum number of points no three on one line) in finite Desarguesian planes. This method can be applied to any plane of order  $2^h$ ,  $h$  a positive integer, i.e., to planes with  $2^h + 1$  points on a line. No general procedure has thus far been obtained for planes with  $p^n + 1$  points on a line,  $p$  an odd prime and  $n$  a positive integer. A design based on a plane with 10 points on a line will be constructed. Further generalizations of this method will be discussed later.

**2. Construction of partially balanced designs based on finite Desarguesian planes with  $2^h + 1$  points on a line.** The total number of points in a plane with  $2^h + 1$  points on a line is  $2^{2h} + 2^h + 1$ . Furthermore it is well known that such planes include ovals consisting of the maximum possible number of points  $2^h + 2$ . The lines of the plane can be classified into two categories with respect to the oval. The first category includes lines having two points of the oval, henceforth called secants. The second category of lines consists of lines not including any point of the oval. The number of lines belonging to each of the two categories is clearly  $(2^{h-1} + 1)(2^h + 1)$  and  $2^{2h-1} - 2^{h-1}$  respectively. Consider now the points of the plane which are not on the oval. Their number is  $2^{2h} - 1$ . Each of them lies on  $2^{h-1} + 1$  secants and  $2^{h-1}$  lines of the second category. This leads to a construction of partially balanced block designs identifying the points with the objects and the lines with the blocks. Each of the two categories of lines taken separately gives rise to a partially balanced block design. The first design will be obtained by calling two objects first associates if the points representing them lie on one secant, second associates otherwise. The second design will be obtained by interchanging the roles of the two categories of lines.

The parameters of the first design are as follows:

$$\begin{aligned} v &= 2^{2h} - 1, & b &= (2^{h-1} + 1)(2^h + 1), \\ r &= 2^{h-1} + 1, & \lambda_1 &= 1, & \lambda_2 &= 0, \\ k &= 2^h - 1, & n_1 &= 2^{2h-1} - 2, & n_2 &= 2^{2h-1}, & n_2 &= 2^{2h-1}, \\ P_1 &= \left( 2^{2h-2} - 3, \begin{matrix} 2^{2h-2} \\ 2^{2h-2} \end{matrix} \right), & P_2 &= \left( 2^{2h-2} - 1, \begin{matrix} 2^{2h-2} - 1 \\ 2^{2h-2} \end{matrix} \right). \end{aligned}$$

The parameters of the second design obtained by identifying the blocks with

Received October 7, 1960; revised May 15, 1961.

the lines not including points of the oval are

$$\begin{aligned} v &= 2^h - 1, & b &= 2^{2h-1} - 2^{h-1}, & r &= 2^h - 2^{h-1}, & k &= 2^h + 1, \\ \lambda_1 &= 1, & \lambda_2 &= 0, & n_1 &= 2^{2h-1}, & n_2 &= 2^{2h-1} - 2, \\ P_1 &= \begin{pmatrix} 2^{2h-2} & 2^{2h-2} - 1 \\ 2^{2h-2} & -1 \end{pmatrix}, & P_2 &= \begin{pmatrix} 2^{2h-2} & 2^{2h-2} \\ 2^{2h-2} & -3 \end{pmatrix}. \end{aligned}$$

Let us illustrate the above described method applying it to the case  $h = 2$ , i.e., to the finite projective plane with 5 points on a line. Take the oval represented by the unruled conic  $x^2 + yz = 0$  and the point of intersection of the tangents to this conic, where  $x, y, z$  are the coordinates of a point in the plane. Let  $\epsilon$  be a primitive element of the corresponding Galois field. The six points of the oval are then:  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ ,  $(1, 1, 1)$ ,  $(1, \epsilon, \epsilon^2)$ ,  $(1, \epsilon^2, \epsilon)$ . The remaining fifteen points of the plane are:  $(0, 1, 1)$ ,  $(0, 1, \epsilon)$ ,  $(0, 1, \epsilon^2)$ ,  $(1, 0, 1)$ ,  $(1, 0, \epsilon)$ ,  $(1, 0, \epsilon^2)$ ,  $(1, 1, 0)$ ,  $(1, 1, \epsilon)$ ,  $(1, 1, \epsilon^2)$ ,  $(1, \epsilon, 0)$ ,  $(1, \epsilon, 1)$ ,  $(1, \epsilon, \epsilon)$ ,  $(1, \epsilon^2, 0)$ ,  $(1, \epsilon^2, 1)$ ,  $(1, \epsilon^2, \epsilon^2)$ . Thus we can exhibit a partially balanced block design with the following parameters:  $v = 15$ ,  $b = 15$ ,  $r = 3$ ,  $k = 3$ ,  $\lambda_1 = 1$ ,  $\lambda_2 = 0$ ,  $n_1 = 6$ ,  $n_2 = 8$ ,  $P_1 = \begin{pmatrix} 1 & 4 \\ 4 & 4 \end{pmatrix}$ ,  $P_2 = \begin{pmatrix} 3 & 3 \\ 3 & 4 \end{pmatrix}$ . If we identify the points with the consecutive numbers 1 through 15 we obtain Table I, the plan of the design in question.

TABLE I  
*The Plan of the Design*

1	2	3	3	5	13	6	9	15
1	4	7	3	9	11	7	8	9
1	12	15	4	5	6	7	10	13
2	8	14	4	11	14	10	11	12
2	6	10	5	8	12	13	14	15

REMARK 1. The described method of construction of partially balanced block designs also gives a method of construction of Hadamard matrices of order  $2^h$ , a method that differs from the one given by Paley [1]. The elements of these matrices can be rearranged so that they have a constant element on the diagonal and are symmetric about the main diagonal. They have further properties which will not be described here. The association matrices of the design [2] yield such matrices. Here is an example of such a matrix for  $h = 2$  based on the partially balanced block design with blocks represented by the second category of lines. The diagonal elements are zeros and the below diagonal part of the matrix is as follows in Table II.

REMARK 2. It seems worthwhile to point out that the method of construction of partially balanced block designs was applied to Desarguesian planes only because of the fact that they do include an oval consisting of  $2^h + 2$  points. If the same would hold with respect to non-Desarguesian planes, then conceivably one could obtain, using the same method, designs that are in general non-isomorphic with the ones already constructed.



TABLE II

1  
 11  
 111  
 1111  
 10100  
 100101  
 1000111  
 11000111  
 010100101  
 0010110100  
 01001001111  
 001101100110  
 000110110111  
 01100100111110

3. Construction of a partially balanced block design based on planes of order  $p^n$ . No general theory is yet available for construction of designs based on projective planes with  $p^n + 1$  points on a line,  $p$  an odd prime,  $n$  a positive integer. An indication of the approach will be given by examining the case  $p = 3$ ,  $n = 2$ . The total number of points in this plane is 91; the oval consists of 10 points. The 81 remaining points can be classified into five categories depending on the number of secants passing through the points. It will be convenient to name the lines including just one point of the oval tangents. Let us denote the number of points in each of the five categories by  $x$ ,  $y$ ,  $z$ ,  $u$ ,  $w$  respectively. The classification is summarized in Table III.

TABLE III

Category	Lines passing through the point			
	Number of points	Secants	Tangents	Lines not including points of the oval
I	$x$	5	—	5
II	$y$	4	2	4
III	$z$	3	4	3
IV	$u$	2	6	2
V	$w$	1	8	1

Clearly  $x + y + z + u + w = 81$ . Further equations are obtained counting the number of intersections of tangents, secants, separately and the number of those intersections of the secants with the tangents that are not points of the oval. This yields the following equations:

$$10x + 6y + 3z + u = 630,$$

$$y + 6z + 15u + 28w = 45,$$

$$2y + 3z + 3u + 2w = 90.$$

It is easy to show that the only possible positive integer solutions of the above equations are  $x = 36$ ,  $y = 45$ . This leads to a partially balanced block design if we identify, e.g., the objects with the  $y$ 's and the blocks with the tangents excluding the points of the oval. The parameters of the design are:  $v = 45$ ,  $b = 10$ ,  $r = 2$ ,  $n_1 = 16$ ,  $n_2 = 28$ ,  $\lambda_1 = 1$ ,  $\lambda_2 = 0$ ,  $k = 9$ ,  $P_1 = \begin{pmatrix} 8 & 7 \\ 21 \end{pmatrix}$ ,  $P_2 = \begin{pmatrix} 4 & 12 \\ 15 \end{pmatrix}$ . The plan of the design is shown in Table IV.

TABLE IV

1	2	3	4	5	6	7	8	9
1	10	11	12	13	14	15	16	17
2	10	18	19	20	21	22	23	24
3	11	18	25	26	27	28	29	30
4	12	19	25	31	32	33	34	35
5	13	20	26	31	36	37	38	39
6	14	21	27	32	36	40	41	42
7	15	22	28	33	37	40	43	44
8	16	23	29	34	38	41	43	45
9	17	24	30	35	39	42	44	45

It may be noticed that this design could be obtained by not using any theory but the method will apply to more complicated cases.

In conclusion I wish to thank R. C. Bose for a stimulating discussion which led to this paper.

## REFERENCES

- [1] R. E. A. C. PALEY, "On orthogonal matrices," *J. Math. Phys.*, Vol. 12 (1933), pp. 311-320.
- [2] R. C. BOSE AND DALE MESNER, "On linear associative algebras corresponding to association schemes of partially balanced block designs," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 21-38.

# ON SOME METHODS OF CONSTRUCTION OF PARTIALLY BALANCED ARRAYS<sup>1</sup>

By I. M. CHAKRAVARTI<sup>2</sup>

*University of North Carolina*

**Summary.** Partially balanced arrays are generalizations of orthogonal arrays. Multifactorial designs derived from partially balanced arrays require a reduced number of assemblies in order to accommodate a given number of factors. For instance, an orthogonal array of strength two, six symbols and four constraints, would require at least  $2.6^2 = 72$  assemblies. This is because there does not exist a pair of mutually orthogonal Latin Squares of order six. But for the same situation, a partially balanced array in 42 assemblies, is constructed in this paper. The method of construction is one of composition which utilizes the existence of a *pairwise partially balanced incomplete block design* and an orthogonal array.

**1. Introduction.** Suppose  $A = (a_{ij})$  is a matrix,  $i = 1, \dots, m, j = 1, \dots, N$  and the elements  $a_{ij}$  of the matrix are symbols  $0, 1, 2, \dots, s - 1$ . Consider the  $s' 1 \times t$  matrices  $X' = (x_1, x_2, \dots, x_t)$  that can be formed by giving different values to the  $x_i$ 's,  $x_i = 0, 1, 2, \dots, s - 1; i = 1, \dots, t$ . Suppose associated with each  $t \times 1$  matrix  $X$  there is a positive integer  $\lambda(x_1, x_2, \dots, x_t)$  which is invariant under permutations of  $(x_1, x_2, \dots, x_t)$ . If, for every  $t$ -rowed submatrix of  $A$ , the  $s' t \times 1$  matrices  $X$  occur as columns  $\lambda(x_1, x_2, \dots, x_t)$  times, then the matrix  $A$  is called a *partially balanced array* of strength  $t$  in  $N$  assemblies,  $m$  constraints (or factors),  $s$  symbols (or levels) and the specified  $\lambda(x_1, x_2, \dots, x_t)$  parameters. When  $\lambda(x_1, x_2, \dots, x_t) = \lambda$  for all  $(x_1, x_2, \dots, x_t)$ , the array is called an orthogonal array.

Orthogonal arrays were defined in [6] and [7] and construction of orthogonal arrays were considered in [1], [2], [3], [6] and [7]. Partially balanced arrays were defined in [5], where their use as multifactorial designs is also discussed.

In this paper, some methods of construction of partially balanced arrays are considered. One of the methods is applicable when  $s = 2$  and derives partially balanced arrays from the well-known  $\lambda - \mu - \nu$  configurations. The other method is an extension of the Bose-Shrikhande [2] method of construction of orthogonal arrays.

**2. An example of a partially balanced array.** Deleting the first three assemblies and the last row from the orthogonal array  $A(18, 7, 3, 2)$  given in [1], one gets a

Received October 24, 1960; revised May 26, 1961.

<sup>1</sup> This research was supported in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

<sup>2</sup> Present address: Statistical Laboratory, Case Institute of Technology, Cleveland 6, Ohio.

partially balanced array of strength two,  $s = 3$  symbols and  $m = 6$  constraints in  $N = 15$  assemblies. This array has the  $\lambda(x_1, x_2)$  parameters

$$\begin{aligned}\lambda(x_1, x_2) &= 2 && \text{if } x_1 \text{ and } x_2 \text{ are unlike,} \\ &= 1 && \text{otherwise.}\end{aligned}$$

The orthogonal array and the derived partially balanced array are given in Tables 1 and 2. The columns of the partially balanced array were a little re-arranged.

TABLE 1  
*Orthogonal Array A(18, 7, 3, 2) assemblies*

Constraints	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
2	0	1	2	0	1	2	1	2	0	2	0	1	1	2	0	2	0	1
3	0	1	2	1	2	0	0	1	2	2	0	1	2	0	1	1	2	0
4	0	1	2	2	0	1	2	0	1	0	1	2	1	2	0	1	2	0
5	0	1	2	1	2	0	2	0	1	1	2	0	0	1	2	2	0	1
6	0	1	2	2	0	1	1	2	0	1	2	0	2	0	1	0	1	2
7	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2

TABLE 2  
*Partially balanced array (15, 6, 3, 2) assemblies*

Constraints	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0	0	0	0	1	1	1	1	1	2	2	2	2	2
2	0	2	1	1	2	0	0	1	2	2	0	0	2	1	1
3	1	0	2	2	1	0	2	0	1	2	1	2	0	0	1
4	1	1	0	2	2	2	0	2	0	1	2	1	0	1	0
5	2	2	1	0	1	1	2	2	0	0	0	1	1	0	2
6	2	1	2	1	0	2	1	0	2	0	1	0	1	2	0

Suppose, an orthogonal array  $A(N, m, s, t)$  of index  $\lambda$  is resolvable into two disjoint arrays. Further, let one of them be a partially balanced array or a *degenerate* partially balanced array (a degenerate array being one which has some but not all  $\lambda(x_1, x_2, \dots, x_t)$  equal to zero), with  $\lambda(x_1, x_2, \dots, x_t) < \lambda$  for all  $t \times 1$  matrices  $X$ . Then the residual array is a partially balanced array with  $\lambda$ -parameters  $\bar{\lambda}(x_1, x_2, \dots, x_t) = \lambda - \lambda(x_1, x_2, \dots, x_t)$ . This provides a basis for the deletion process of deriving a partially balanced array from an orthogonal array.

### 3. Construction of partially balanced arrays for $s = 2$ from $\lambda - \mu - \nu$ configurations.

DEFINITION. A  $\lambda - \mu - \nu$  configuration of  $m$  elements is defined [4] as the configuration of  $m$  elements taken  $\nu$  at a time so that each set of  $\mu$  elements shall occur together in just  $\lambda$  of the sets.

Suppose there are  $N_0$  sets of  $\nu$  elements each in the configuration. Let  $N_t$  denote the number of sets each containing a fixed subset of  $t$  elements. Then it is easily

seen that

$$(3.1) \quad N_t = \lambda \binom{m-t}{\mu-t} / \binom{\nu-t}{\mu-t} \quad t = 0, 1, 2, \dots, \mu.$$

Consider the matrix  $A = (a_{ij})$  of order  $m \times N_0$  derived from a  $\lambda - \mu - \nu$  configuration of  $m$  elements in  $N_0$  sets in the following manner: Let  $\alpha_1, \alpha_2, \dots, \alpha_m$  denote the  $m$  elements and  $s_1, s_2, \dots, s_{N_0}$  denote the  $N_0$  sets of the configuration. Then let

$$\begin{aligned} a_{ij} &= 1 && \text{if } \alpha_i \text{ occurs in the set } s_j \\ &= 0 && \text{otherwise.} \end{aligned}$$

Consider a  $\mu$ -rowed submatrix of  $A$  with elements  $a_{ij}$  as defined above. Amongst the  $N_0$  columns of the submatrix, a column matrix  $X_{\mu,1}$  where its transpose  $X'_{1,\mu} = (x_1, x_2, \dots, x_\mu)$ ,  $x_i = 0$  or  $1$ ,  $i = 1, \dots, \mu$  occurs  $\lambda(x_1, x_2, \dots, x_\mu)$  times. Specifically, let  $x_i = 1$  for  $i = 1, \dots, r$  and let  $x_i = 0$  for  $i = r+1, \dots, \mu$  in  $X$ . Then it is easy to show that for such an  $X$

$$\begin{aligned} (3.2) \quad \lambda(x_1, x_2, \dots, x_\mu) &= N_r - \binom{\mu-r}{1} N_{r+1} + \binom{\mu-r}{2} N_{r+2} - \dots \\ &= (-1)^{\mu-r} \Delta^{\mu-r} N_r \end{aligned}$$

where  $\Delta$  stands for the symbol of finite difference,

$$\Delta N_r = N_{r+1} - N_r.$$

Value of  $\lambda(x_1, x_2, \dots, x_\mu)$  depends only on the count  $r$  of unities in its argument and hence it is invariant under permutation of its arguments.

Now provided  $\lambda(x_1, x_2, \dots, x_\mu) > 0$  for all  $s^\mu$  sets of  $X$ , we have

**THEOREM 2.1.** *The existence of a  $\lambda - \mu - \nu$  of  $m$  elements with  $\lambda(x_1, x_2, \dots, x_\mu)$  all positive, implies the existence of a partially balanced array of strength  $\mu$  with parameters  $s = 2$  and  $\lambda(x_1, x_2, \dots, x_\mu)$  as defined in (3.2).*

Well known examples of  $\lambda - \mu - \nu$  configurations are the triple systems, quadruple systems, etc., which are defined in [4].

#### 4. An extension of the Bose-Shrikhande method of construction of orthogonal arrays and its use in the construction of partially balanced arrays.

**DEFINITION.** A pairwise partially balanced design with parameters

$$(\nu, k_1, k_2, \dots, k_m; b_1, b_2, \dots, b_m; \lambda_1, \lambda_2, \dots, \lambda_t; n_1, n_2, \dots, n_t)$$

is defined as an arrangement of  $\nu$  varieties in blocks of  $m$  different sizes  $k_1, k_2, \dots, k_m$ , there being  $b_i$  blocks of size  $k_i$ ,  $\sum_{i=1}^m b_i = b$ , satisfying the following conditions:

- (i) No block contains a single variety more than once.
- (ii) With respect to any variety, the remaining  $\nu - 1$  varieties fall into  $t$  categories, there being  $n_i$  varieties in the  $i$ th category, called the  $i$ th associates of the variety;  $\sum_{i=1}^t n_i = \nu - 1$ .

(iii) Two varieties which are  $i$ th associates, occur together in  $\lambda_i$  blocks,  $i = 1, \dots, t$ .

Then the following relations among the parameters hold,

$$(4.1) \quad \sum_{i=1}^m b_i k_i (k_i - 1) = \sum_{i=1}^t n_i \nu \lambda_i = \nu \sum_{i=1}^t n_i \lambda_i.$$

Suppose there exist the orthogonal arrays

$$A_i (\lambda k_i^2, q_i, k_i, 2) \quad i = 1, \dots, m$$

of strength two and index  $\lambda$  and in  $k_i$  symbols. Consider the pairwise partially balanced design defined earlier. There are  $b_i$  blocks each of size  $k_i$ . These  $b_i$  blocks provide  $b_i$  sets of  $k_i$  symbols each. Using each set of  $k_i$  symbols once in the orthogonal array  $A_i$ , one gets  $b_i$  such orthogonal arrays. If all such orthogonal arrays are arranged side by side, then one gets a matrix  $A$  with number of columns  $N = \lambda \sum_{i=1}^m b_i k_i^2$  and number of rows  $q = \min(q_1, q_2, \dots, q_m)$ . In the columns of any two-rowed submatrix of matrix  $A$ , every ordered pair  $(t_u, t_v)$  of two distinct symbols of varieties which are  $i$ th associates will occur  $\lambda \lambda_i$  times and every ordered pair  $(t_j, t_j)$  of two like symbols occur  $\lambda r_j$  times, if the variety  $t_j$  occurs in  $r_j$  blocks of the pairwise partially balanced design. Hence we have

**THEOREM 4.1.** *The existence of a pairwise partially balanced design with parameters  $(\nu; k_1, k_2, \dots, k_m; b_1, b_2, \dots, b_m; \lambda_1, \lambda_2, \dots, \lambda_t; n_1, n_2, \dots, n_t)$  and of the orthogonal arrays  $A_i(\lambda k_i^2, q_i, k_i, 2)$   $i = 1, \dots, m$ , imply the existence of the partially balanced array of strength two in  $\nu$  symbols and  $q = \min(q_1, q_2, \dots, q_m)$  constraints and  $\lambda(x_1, x_2) = \lambda \lambda_i$ , where  $x_1, x_2$  stand for two varieties which are  $i$ th associates and  $\lambda(x, x) = \lambda r_j$ , and where the variety  $x$  occurs  $r_j$  times in the pairwise partially balanced design.*

As an illustration, a partially balanced array which has been constructed using the method described above, is given below. This is a partially balanced array in  $\nu = 6$  symbols,  $N = 48$  assemblies,  $m = 5$  constraints and

$$\begin{aligned} \lambda(x_1, x_2) &= 2 && \text{if } (x_1, x_2) \text{ are first associates} \\ &= 1 && \text{if } (x_1, x_2) \text{ are second associates} \\ &= 2 && \text{if } x_1 \text{ and } x_2 \text{ are like,} \end{aligned}$$

where  $x_i, i = 1, \dots, 6$  are the variety symbols. In constructing this array, the partially balanced design

$$(\nu = 6, r = 2, b = 3, k = 4, n_1 = 1, n_2 = 4, \lambda_1 = 2, \lambda_2 = 1)$$

in three blocks

$$\begin{array}{cccc} x_1 & x_4 & x_2 & x_5 \\ x_2 & x_5 & x_3 & x_6 \\ x_3 & x_6 & x_1 & x_4 \end{array}$$

and the orthogonal array  $A(16, 5, 4, 2)$  have been used.



TABLE 3  
Orthogonal Array A(16, 5, 4, 2)

R	0	0	0	0	1	1	1	1	t	t	t	t	t <sup>2</sup>	t <sup>2</sup>	t <sup>2</sup>	t <sup>2</sup>
C	0	1	t	t <sup>2</sup>	0	1	t	t <sup>2</sup>	0	1	t	t <sup>2</sup>	0	1	t	t <sup>2</sup>
L <sub>1</sub>	0	1	t	t <sup>2</sup>	1	0	t <sup>2</sup>	t	t	t <sup>2</sup>	0	1	t <sup>2</sup>	t	1	0
L <sub>2</sub>	0	1	t	t <sup>2</sup>	t	t <sup>2</sup>	0	1	t <sup>2</sup>	t	1	0	1	0	t <sup>2</sup>	t
L <sub>3</sub>	0	1	t	t <sup>2</sup>	t <sup>2</sup>	t	1	0	1	0	t <sup>2</sup>	t	t	t <sup>2</sup>	0	1

Making successively the identifications

$$\begin{array}{lll}
 (1) & (2) & (3) \\
 1 = x_1 & 1 = x_2 & 1 = x_1 \\
 t = x_2 & t = x_3 & t = x_3 \\
 t^2 = x_4 & t^2 = x_5 & t^2 = x_4 \\
 0 = x_5 & 0 = x_6 & 0 = x_6
 \end{array}$$

and using them on the above array in place of  $(0, 1, t, t^2)$ , one gets three arrays,  $A_1$ ,  $A_2$  and  $A_3$ . Then the array  $A_0 = [A_1 A_2 A_3]$  is the desired partially balanced array in 6 symbols and 48 assemblies. Let  $A^*$  denote the array derived from  $A$  by truncating the first row and the first four columns (as indicated by the horizontal and vertical lines). Then the arrays  $A_1^*$ ,  $A_2^*$  and  $A_3^*$  are obtained from  $A^*$  using the three identifications of variety-symbols given above. Let  $E$  denote the array

$$E: \begin{bmatrix} x_1 & x_2 & \cdots & \cdots & x_6 \\ x_1 & x_2 & \cdots & \cdots & x_6 \\ x_1 & x_2 & \cdots & \cdots & x_6 \\ x_1 & x_2 & \cdots & \cdots & x_6 \end{bmatrix}$$

Then the array  $A_0^* = [E A_1^* A_2^* A_3^*]$  is a partially balanced array in  $v = 6$  symbols,  $N = 42$  assemblies,  $m = 4$  constraints and  $\lambda(x, x) = 1$ ,  $\lambda(x_i, x_j) = 2$  if  $x_i$  and  $x_j$  are first associates and  $\lambda(x_i, x_j) = 1$  if they are second associates.

**5. Acknowledgment.** Thanks are due to Professor R. C. Bose for kindly going through the manuscript and for his helpful comments.

#### REFERENCES

- [1] BOSE, R. C. AND BUSH, K. A., "Orthogonal arrays of strength two and three," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 508-524.
- [2] BOSE, R. C. AND SHRIKHANDE, S. S., "On the composition of balanced incomplete block designs," *Canad. J. Math.*, Vol. 12 (1960), pp. 177-188.
- [3] BUSH, K. A., "Orthogonal arrays of index unity," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 426-434.
- [4] CARMICHAEL, R. D., *Introduction to the Theory of Groups of Finite Order*, Dover Publications, New York, 1937.
- [5] CHAKRAVARTI, I. M., "Fractional replication in asymmetrical factorial designs and partially balanced arrays," *Sankhyā*, Vol. 17 (1956), pp. 143-164.
- [6] RAO, C. R., "On hypercubes of strength  $d$  and a system of confounding in factorial experiments," *Bull. Cal. Math. Soc.* Vol. 38 (1946), pp. 67-78.
- [7] RAO, C. R., "Factorial experiments derivable from combinatorial arrangements of arrays," *J. Roy. Stat. Soc. (Suppl.)*, Vol. 9 (1947), pp. 128-139.

## SOME FURTHER DESIGNS OF TYPE O:PP

BY G. H. FREEMAN

*West African Cocoa Research Institute, Ibadan, Nigeria*

**0. Summary.** A new method of deriving designs of type O:PP is described. The method gives rise to some designs previously obtained by other methods, and also to some entirely new designs. These new designs are described in detail and a worked example given.

**1. Introduction.** Experimental designs with three non-interacting classifications are most likely to be of use in two cases. The first is when the trial area has two physical configurations, such as rows and columns in a field; the second is when a new set of treatments is added to an existing block design and this set is, by its nature, unlikely to interact with the previous treatments. In either case, orthogonal designs, if available, are the best, but frequently the numbers of treatments and other classifications make complete orthogonality impossible. The problems of experimental design are the same in both cases, and suitable experimental designs using total or partial balance were considered by Hoblyn, *et al.* [4]. In their notation, if an experimental design has three classifications, rows, columns and treatments, such that the rows and columns are orthogonal to each other and the treatments are partially balanced with respect to both rows and columns, then the design is said to be of type O:PP.

Designs of type O:PP were discussed in more detail by Freeman [1], who stated that the only practicable designs are those where the designs of type  $P$  have two associate-classes only and these two associate-classes are the same for each  $P$ . It appears, however, that useful O:PP designs with two associate-classes can sometimes be derived from two designs of type  $P$  without these restrictions. One possibility is that the two designs of type  $P$  have two associate-classes each but that these are not the same for rows and columns: another is that the two designs of type  $P$  each have three associate-classes. Since designs with two associate-classes can be regarded as special cases of those with three associate-classes in which some of the parameters are equal, the first of these possibilities can perhaps be regarded as included within the second; nevertheless, it is probably better to retain the distinction.

In order that the resultant O:PP design shall be analysable by the methods previously given [1], the two designs of type  $P$  have to satisfy certain conditions. In the O:PP design let there be  $n$  replicates of  $t$  treatments on  $r$  rows and  $c$  columns such that each treatment occurs either  $f$  or  $f + 1$  times in rows and either  $g$  or  $g + 1$  times in columns. Let the designs of type  $P$  for rows and columns have three associate-classes, the same for each classification, with  $n_i$  members in the  $i$ th class. Amongst the extra occurrences let  $i$ th associates concur  $\lambda_i$  times

Received February 17, 1961; revised July 11, 1961.

in rows and  $\mu_i$  times in columns, and write  $\nu_i = r\lambda_i + c\mu_i$ . Then, for the design of type O:PP to have only two associate-classes, two of  $\nu_1, \nu_2$  and  $\nu_3$  must be equal. If  $\nu_i = \nu_j \neq \nu_k$ , where  $i, j, k$  are 1, 2, 3 in some order, then, in the O:PP design, any one treatment has  $n_i + n_j$  associates of one kind and  $n_k$  of the other. Which of these are called first associates and which second is largely a matter of convenience, but the same convention as for partially balanced designs should be used when possible. It is also necessary for the parameters of the second kind in the O:PP design to satisfy the usual equations for a partially balanced design, but these parameters cannot in general be derived directly from those of the two separate designs of type *P*. When the two designs of type *P* each have only two associate-classes, these not being the same, this is represented by  $\lambda_3 = \lambda_1, \mu_3 = \mu_2$ .

**2. Possible designs.** Freeman [3] gave a catalogue of useful designs derived in the orthodox manner, "useful" designs being those having more than two replicates or treatments, not more than 30 replicates, treatments, rows or columns, and not more than 150 plots in all. The method of obtaining designs described here does not lead to many useful new designs, while there are some which have the same parameters as those derived by the old method. When these are mentioned in this section they are numbered as in Freeman [3]. Thus, there is a design with 8 replicates of 18 treatments on 12 rows and columns,  $n_1 = 1, n_2 = 8, n_3 = 8, \lambda_1 = 8, \lambda_2 = 6, \lambda_3 = 4, \mu_1 = 8, \mu_2 = 4, \mu_3 = 6$ . Hence  $\nu_1 = 192, \nu_2 = \nu_3 = 120$ , and so the design has the same parameters as *SS VI 1*.

There are two designs which may be of more use, and one has already been used to design a trial in the field. Both use the principle of having singular group-divisible designs for each of rows and columns, the resultant O:PP design being of Latin square type. All useful Latin square designs so far discovered, whether obtained by this method or the orthodox one, have equal numbers of rows and columns. They are shown in Table I, which thus repeats some of the information given previously [3]. The horizontal line through the middle of Table I separates designs derived by the two methods. Above the line  $\nu_1 = r\lambda_1 + c\mu_1, \nu_2 = r\lambda_2 + c\mu_2$ , as usual, while below the line  $\nu_1 = r\lambda_1 + c\mu_1 = r\lambda_2 + c\mu_2, \nu_2 = r\lambda_3 + c\mu_3$ , that is,  $\nu_2 = r\lambda_1 + c\mu_2$ .

TABLE I  
*Useful designs in family LL*

Design	<i>m</i>	<i>w</i>	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\mu_1$	$\mu_2$	$\mu_3$	$\nu_1$	$\nu_2$	Rep	Tr	Row	Col
LL 1	2	3	3	2	—	3	2	—	36	24	4	9	6	6
LL 2	2	3	3	2	—	2	3	—	30	30	4	9	6	6
LL 3	3	4	6	7	—	6	7	—	144	168	9	16	12	12
LL 4	3	4	6	9	6	9	6	6	180	120	9	16	12	12
LL 5	2	5	1	4	1	4	1	1	50	20	4	25	10	10

The most general Latin square design with equal numbers of rows and columns has  $m^2$  replicates of  $w^2$  treatments on  $mw$  rows and columns. When the designs of type  $P$  are singular group-divisible a given treatment in them has  $w - 1$  first associates and  $w(w - 1)$  second associates. Further, if  $fm < w < (f + 1)m$ ,  $\lambda_1 = m^2 - fmw$ , and so

$$\lambda_2 = \frac{(m^2 - fmw)(m - fw - 1)}{w - 1},$$

in order to satisfy the usual constraint on parameters of the first kind.  $\lambda_2$  has to be integral, and this condition imposes some limitation on the possible values of  $m$  and  $w$ . All singular group-divisible designs are derived from balanced incomplete block designs by replacing one treatment of the balanced design by a group of treatments,  $\lambda_2$  in the partially balanced design equalling  $\lambda$  in the balanced design; it is easy to construct corresponding O:PP designs, which are very numerous. Thus, any balanced incomplete block design with  $m$  replicates and plots per block and  $w$  treatments and blocks gives rise to a design with the required parameters. The smallest resultant O:PP design has 4 replicates of 9 treatments on 6 rows and columns, but this has the same parameters as *LL* 1. The next in this series has 9 replicates of 16 treatments on 12 rows and columns, and is shown as *LL* 4 in Table I. There is a group-divisible O:PP design, *SSI* 15, with 9 replicates of 16 treatments, in 4 groups of 4, on 12 rows and columns, but this has all other parameters different. Also, the design shown as *LL* 3, with the same values of  $m$  and  $w$ , which is derived by orthodox methods, is new, although it should have been included in the 1958 paper [3].

Another set of designs arises from the balanced incomplete block designs with  $m^2$  replicates of  $w$  treatments on  $mw$  blocks of  $m$  plots each. The smallest design in this series has  $m = 2$ ,  $w = 5$ , and gives rise to *LL* 5. This is particularly noteworthy in that no other O:PP design is possible with these numbers of rows, columns, treatments and replicates. It seems at first sight that there should be a singular group-divisible O:PP design using the same designs of type  $P$  for rows and columns, but this design is excluded by Theorem 1 [2]. This same Theorem excludes singular group-divisible designs with the same parameters as other Latin square designs in this series, for example, those with  $m = 3$ ,  $w = 10$ , and  $m = 3$ ,  $w = 19$ .

When a  $w \times w$  factorial experiment is to be laid out in rows and columns, a Latin square O:PP design derived from two singular designs of type  $P$  may be particularly suitable. The main effects of the two factors can be associated with the rows and the columns, and the corresponding sums of squares in the analysis take a fairly simple form. Competing designs for the same numbers of treatments and replicates will include lattice squares, but the two designs may well require different numbers of rows and columns, which are often pre-determined. Further, two error variances have to be calculated in a lattice square, and only one in an O:PP design. The example which follows illustrates the method of analysis for a general O:PP design, with the modification required for a  $w \times w$  factorial experiment.

**3. Example.** The design *LL* 5 has been used for a trial on yams (*Dioscorea* sp.) conducted by Mr. E. F. I. Baker, Research Division, Ministry of Agriculture and Natural Resources, Western Region of Nigeria. Yams are grown from setts, pieces of root tuber with adventitious buds, and the purpose of this trial was to find the effect of planting setts of different weights at varying populations per acre. A  $5 \times 5$  factorial system was used, the levels of one factor being sett weights per acre and of the other factor populations per acre. Representing the 25 treatment combinations by

<i>A</i>	<i>F</i>	<i>K</i>	<i>P</i>	<i>U</i>
<i>B</i>	<i>G</i>	<i>L</i>	<i>Q</i>	<i>V</i>
<i>C</i>	<i>H</i>	<i>M</i>	<i>R</i>	<i>W</i> ,
<i>D</i>	<i>I</i>	<i>N</i>	<i>S</i>	<i>X</i>
<i>E</i>	<i>J</i>	<i>O</i>	<i>T</i>	<i>Y</i>

the rows were lettered *a, b, c, d, e* and represented sett weights per acre, and the columns, numbered 1, 2, 3, 4, 5, were populations per acre. Furthermore, the treatments down the leading diagonal, *A, G, M, S, Y* all had the same individual sett weight, though varying in population and weight per acre. The lay-out of the O:PP design in the field, after randomisation of rows and columns, was that shown in Table II.

TABLE II

<i>E</i>	<i>T</i>	<i>J</i>	<i>D</i>	<i>S</i>	<i>Y</i>	<i>X</i>	<i>N</i>	<i>O</i>	<i>I</i>	<i>d e</i>
<i>X</i>	<i>N</i>	<i>C</i>	<i>R</i>	<i>I</i>	<i>S</i>	<i>W</i>	<i>M</i>	<i>D</i>	<i>H</i>	<i>c d</i>
<i>D</i>	<i>S</i>	<i>I</i>	<i>A</i>	<i>F</i>	<i>P</i>	<i>U</i>	<i>X</i>	<i>N</i>	<i>K</i>	<i>a d</i>
<i>V</i>	<i>Q</i>	<i>D</i>	<i>S</i>	<i>G</i>	<i>X</i>	<i>I</i>	<i>L</i>	<i>B</i>	<i>N</i>	<i>b d</i>
<i>U</i>	<i>R</i>	<i>F</i>	<i>C</i>	<i>P</i>	<i>W</i>	<i>H</i>	<i>K</i>	<i>A</i>	<i>M</i>	<i>a c</i>
<i>C</i>	<i>L</i>	<i>B</i>	<i>Q</i>	<i>H</i>	<i>R</i>	<i>V</i>	<i>W</i>	<i>M</i>	<i>G</i>	<i>b c</i>
<i>B</i>	<i>K</i>	<i>A</i>	<i>P</i>	<i>Q</i>	<i>V</i>	<i>G</i>	<i>U</i>	<i>L</i>	<i>F</i>	<i>a b</i>
<i>A</i>	<i>P</i>	<i>E</i>	<i>T</i>	<i>J</i>	<i>U</i>	<i>F</i>	<i>Y</i>	<i>K</i>	<i>O</i>	<i>a e</i>
<i>Y</i>	<i>O</i>	<i>G</i>	<i>B</i>	<i>T</i>	<i>Q</i>	<i>J</i>	<i>V</i>	<i>E</i>	<i>L</i>	<i>b e</i>
<i>W</i>	<i>M</i>	<i>H</i>	<i>E</i>	<i>R</i>	<i>T</i>	<i>Y</i>	<i>O</i>	<i>C</i>	<i>J</i>	<i>c e</i>
1 5	3 4	1 2	1 4	2 4	4 5	2 5	3 5	1 3	2 3	

It is seen that all sett weights per acre in two populations were arranged in each column, these being shown as 1 5, 3 4, etc., while all populations with two sett weights per acre occurred in each row, these being *d e, c d*, etc. First associates of any treatment were thus those treatments with either the same population or the same sett weight per acre.

The analysis follows the lines previously given [1], the notation used being the same. Thus, in the O:PP design,

$$p_{ij}^1 = \begin{pmatrix} 3 & 4 \\ 4 & 12 \end{pmatrix}, \quad p_{ij}^2 = \begin{pmatrix} 2 & 6 \\ 6 & 9 \end{pmatrix}, \quad N = n(cr - r - c) = 320.$$

If treatment, row and column totals are represented by *D, B* and *C* respec-

tively,  $G$  being the grand total, the treatment total for  $A$  adjusted for rows and columns is  $P_A$ , where

$$P_A = 100T_A - 10(B_2 + B_5 + B_7 + B_8) - 10(C_1 + C_3 + C_4 + C_9) + 4G, \text{ etc.}$$

Further,  $A_{12} = 340$ ,  $B_{12} = -30$ ,  $A_{22} = -180$ ,  $B_{22} = 310$ ,  $\Delta = 100000$ . Then, the treatment parameter for  $A$  is  $\delta_A = (31P_A + 3\sum P_{A1})/10000$ , where  $\sum P_{A1}$  is the sum of the  $P$ 's for the first associates of  $A$ .

The treatment sum of squares is  $\sum \delta P/100$ , and row and column sums of squares are unadjusted.

The variance of the difference between the means of two treatments that are  $i$ th associates is obtained by multiplying the error variance by  $\psi_i^2$ , where  $\psi_1^2 = \frac{1}{25}$ ,  $\psi_2^2 = \frac{3}{10}$ .

In this trial the main comparison of importance was between first associates, these being particular levels of one factor for a given level of the other. The most important comparison for second associates was that among the treatments  $A$ ,  $G$ ,  $M$ ,  $S$ ,  $Y$  with the same individual sett weight. It was also necessary to consider the main effects of populations and sett weights per acre. The sum of squares for the main effect of populations has the simple form

$$\sum (P_A + P_B + P_C + P_D + P_E)^2/125000$$

minus the correction factor, and similarly for sett weights. The multiplying factor for the error variance when comparing the means of any two levels of one factor, taken over all levels of the other, is  $\frac{4}{25}$ .

#### REFERENCES

- [1] G. H. FREEMAN, "Some experimental designs of use in changing from one set of treatments to another, Part I," *J. Roy. Stat. Soc.*, Ser. B, Vol. 19 (1957), pp. 154-162.
- [2] G. H. FREEMAN, "Some experimental designs of use in changing from one set of treatments to another, Part II: Existence of the designs," *J. Roy. Stat. Soc.*, Ser. B, Vol. 19 (1957), pp. 163-165.
- [3] G. H. FREEMAN, "Families of designs for two successive experiments," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1063-1078.
- [4] T. N. HOBLYN, S. C. PEARCE AND G. H. FREEMAN, "Some considerations in the design of successive experiments in fruit plantations," *Biometrics*, Vol. 10 (1954), pp. 503-515.



## SUFFICIENCY IN THE UNDOMINATED CASE<sup>1</sup>

BY D. L. BURKHOLDER

*University of Illinois*

**1. Introduction and summary.** In this paper the concept of statistical sufficiency is studied within a general probability setting. It is not assumed that the family of probability measures is dominated. That is, it is not assumed that there is a  $\sigma$ -finite measure  $\mu$  such that each probability measure in the family is absolutely continuous with respect to  $\mu$ . In the dominated case, the theory of sufficiency has received a thorough-going and elegant treatment by Halmos and Savage [6], Bahadur [2], and others. Although many families of probability measures of importance for statistical work are dominated, many others are not. Nonparametric statistical work, especially, abounds with undominated families. It seems appropriate, therefore, to see what can be learned about sufficiency in the undominated case.

Let  $X$  be a set,  $\mathbf{A}$  a  $\sigma$ -field of subsets of  $X$ , and  $P$  a family of probability measures  $p$  on  $\mathbf{A}$ . The probability structure  $(X, \mathbf{A}, P)$  is to be kept in mind throughout the paper and is unrestricted except where specifically stated to the contrary. Any subfield ( $=$  sub- $\sigma$ -field) entering the discussion is implicitly assumed to be a subfield of  $\mathbf{A}$ . If  $H$  is a collection of subfields, let  $\vee H$  denote the smallest  $\sigma$ -field containing each member of  $H$ . If  $\mathbf{A}_1, \mathbf{A}_2, \dots$  are subfields, write  $\mathbf{A}_1 \vee \mathbf{A}_2$  for  $\vee \{\mathbf{A}_1, \mathbf{A}_2\}$ ,  $\bigvee_{n=1}^{\infty} \mathbf{A}_n$  for  $\vee \{\mathbf{A}_1, \mathbf{A}_2, \dots\}$ , and so forth. A set  $N$  is  $P$ -null if  $N$  is  $p$ -null for each  $p$  in  $P$ , that is, if  $N$  is in  $\mathbf{A}$  and  $p(N) = 0$ ,  $p \in P$ . If  $f$  and  $g$  are  $\mathbf{A}$ -measurable functions, write  $f = g[p]$  if the set  $\{x \mid f(x) \neq g(x)\}$  is  $p$ -null and write  $f = g[P]$  if this set is  $P$ -null. Let  $\mathbf{N}$  be the smallest  $\sigma$ -field containing the  $P$ -null sets. If  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are subfields, write  $\mathbf{A}_1 \subset \mathbf{A}_2[P]$  if  $\mathbf{A}_1 \subset \mathbf{A}_2 \vee \mathbf{N}$ , and so forth. A subfield  $\mathbf{B}$  is sufficient if, for each bounded  $\mathbf{A}$ -measurable function  $f$ , there is a  $\mathbf{B}$ -measurable function  $g$  such that  $\int_B f dp = \int_B g dp$ ,  $B \in \mathbf{B}$ ,  $p \in P$ , that is, such that  $g = E_p(f \mid \mathbf{B})[p]$ ,  $p \in P$ . Equivalent definitions are obtained if "bounded  $\mathbf{A}$ -measurable function" is replaced by " $\mathbf{A}$ -measurable characteristic function" or by " $P$ -integrable function." Of course,  $f$  is  $P$ -integrable if  $f$  is  $\mathbf{A}$ -measurable and  $\int_X |f| dp$  is finite for each  $p$  in  $P$ . A subfield  $\mathbf{B}$  is separable if it contains a countable subcollection such that  $\mathbf{B}$  is the smallest  $\sigma$ -field containing the subcollection.

In Section 2, we give an example of a nonsufficient subfield containing a sufficient subfield. This solves a problem posed by Bahadur (Problem 1 on page 441 of [2]). In fact, we show that often the collection of such nonsufficient subfields is much larger than the collection of sufficient subfields. Analogous results hold for statistics. Some of these and later results depend on Theorem 1 which

Received September 9, 1960.

<sup>1</sup> This research was supported by the National Science Foundation under Grant No. G11382.

gives a necessary condition for a subfield to be sufficient in the case that  $\mathbf{A}$  is separable.

Let  $\mathbf{A}_1, \mathbf{A}_2, \dots$  be a sequence of sufficient subfields. Are the subfields  $\mathbf{A}_1 \cap \mathbf{A}_2, \mathbf{A}_1 \vee \mathbf{A}_2, \bigcap_{n=1}^{\infty} \mathbf{A}_n$ , and  $\bigvee_{n=1}^{\infty} \mathbf{A}_n$ , necessarily sufficient? This question is investigated in Sections 3 and 4. Using martingale theory, we show that, if the sequence is decreasing (increasing), then  $\bigcap_{n=1}^{\infty} \mathbf{A}_n (\bigvee_{n=1}^{\infty} \mathbf{A}_n)$  is sufficient. If the sequence is not necessarily monotone, it is still possible to show that  $\mathbf{A}_1 \cap \mathbf{A}_2$  and  $\bigcap_{n=1}^{\infty} \mathbf{A}_n$  are sufficient under a small extra assumption involving  $\mathbf{N}$ . This result rests on a theorem proved in [3] regarding iterates of conditional expectation operators. One consequence of this result is of interest in connection with the theory of minimal sufficient subfields. It is not necessarily true that  $\mathbf{A}_1 \vee \mathbf{A}_2$  is sufficient. This is shown in Example 4. Conditions under which  $\mathbf{A}_1 \vee \mathbf{A}_2$  is sufficient are examined.

The main result of Section 5 is related to Theorem 1 and indicates that if  $\mathbf{A}$  is separable then each sufficient subfield is essentially equal to one of a very special type.

**2. On a problem of Bahadur.** In [2], Bahadur proves that if the family  $P$  of probability measures on  $\mathbf{A}$  is dominated, then a subfield of  $\mathbf{A}$  containing a sufficient subfield is sufficient, and lists as an unsolved problem the question of whether this is true in general. That this is not true in general we now show by an example.

**EXAMPLE 1.** Let  $X$  be the set of real numbers,  $\mathbf{A}$  the collection of Borel subsets of  $X$ , and  $P$  the set of probability measures  $p$  on  $\mathbf{A}$  satisfying  $p(A) = p(-A)$  for  $A$  in  $\mathbf{A}$ . Here, if  $S \subset X$  then  $-S$  is the set  $\{x \mid -x \in S\}$ . Let  $\mathbf{A}_0 = \{A \mid A \in \mathbf{A}, A = -A\}$ . Clearly,  $\mathbf{A}_0$  is a subfield of  $\mathbf{A}$  and if  $f$  is a bounded  $\mathbf{A}$ -measurable function then  $2g(x) = f(x) + f(-x)$  defines an  $\mathbf{A}_0$ -measurable function  $g$  satisfying  $\int_A f dp = \int_A g dp$ ,  $A \in \mathbf{A}_0$ ,  $p \in P$ . Hence,  $\mathbf{A}_0$  is sufficient.

Suppose that  $S$  is a subset of  $X$  satisfying  $0 \in S$  and  $S = -S$ . Let

$$(1) \quad \mathbf{B} = \{A \cup A_0 \mid A \subset S, A \in \mathbf{A}, A_0 \in \mathbf{A}_0\}.$$

Clearly,  $\mathbf{B}$  satisfies  $\mathbf{A}_0 \subset \mathbf{B} \subset \mathbf{A}$ . We now show that  $\mathbf{B}$  is a  $\sigma$ -field. It is obvious that the union of a countable family of sets in  $\mathbf{B}$  is in  $\mathbf{B}$ . Let  $B \in \mathbf{B}$ . Then there are sets  $A$  and  $A_0$  satisfying  $B = A \cup A_0$ ,  $A \subset S$ ,  $A \in \mathbf{A}$ ,  $A_0 \in \mathbf{A}_0$ . Let  $C_0 = (-A) \cup A$  and  $C = C_0 - A$ . Since  $S = -S$ ,  $C_0 \subset S$  and therefore  $C \subset S$ . Using primes to denote complements we have that  $B' = A' \cap A'_0 = (C \cup C'_0) \cap A'_0 = (C \cap A'_0) \cup (C'_0 \cap A'_0)$  which is the union of a subset of  $S$  in  $\mathbf{A}$  and a set in  $\mathbf{A}_0$ . Therefore,  $B' \in \mathbf{B}$  and  $\mathbf{B}$  is a  $\sigma$ -field.

Suppose that  $\mathbf{B}$  is a sufficient subfield. Let  $f$  be a bounded  $\mathbf{A}$ -measurable function. Then, since  $\mathbf{B}$  is sufficient, there is a  $\mathbf{B}$ -measurable function  $g$  satisfying

$$(2) \quad \int_B f dp = \int_B g dp, \quad B \in \mathbf{B}, p \in P.$$

Let  $x \in S$ . Then  $\{x\} \in \mathbf{B}$  and letting  $B = \{x\}$  in (2) gives

$$f(x)p(x) = g(x)p(x), \quad p \in P,$$

where we write  $p(x)$  for  $p(\{x\})$ . Let  $x \in X - S$ . Then  $\{x, -x\} \in \mathbf{B}$  but neither  $\{x\}$  nor  $\{-x\}$  belongs to  $\mathbf{B}$ . Accordingly,  $g(x) = g(-x)$ , since  $g$  is  $\mathbf{B}$ -measurable. Letting  $B = \{x, -x\}$  in (2) and using the fact that  $p(x) = p(-x)$  gives

$$[f(x) + f(-x)]p(x) = 2g(x)p(x), \quad p \in P.$$

If  $x \in X$ , then there is a  $p \in P$  such that  $p(x) > 0$ . Therefore, we have that

$$\begin{aligned} (3) \quad g(x) &= f(x) && \text{if } x \in S, \\ &= \frac{1}{2}[f(x) + f(-x)] && \text{if } x \in X - S. \end{aligned}$$

Let  $f(x) = -1$  if  $x < 0$ ,  $= 1$  if  $x \geq 0$ . Then  $f$  is  $\mathbf{A}$ -measurable and the function  $g$  of (3) is  $\mathbf{B}$ -measurable and satisfies  $g(x) \neq 0$  if  $x \in S$ ,  $= 0$  if  $x \in X - S$ . Thus,  $S = X - g^{-1}(\{0\})$  is in  $\mathbf{B}$ .

Now choose  $S$  to be a subset of  $X$  satisfying  $0 \in S$ ,  $S = -S$ , and  $S \notin \mathbf{A}$ . Such a set exists, of course. Then, if  $\mathbf{B}$  is defined by (1), we see that  $\mathbf{B}$  cannot be sufficient by the result of the above paragraph, for  $S$  does not belong to  $\mathbf{A}$  and therefore does not belong to  $\mathbf{B}$ .

In summary, a subfield can contain a sufficient subfield and yet not be sufficient. We now prove several results which indicate that the probability structure examined in our example is by no means unusual in this respect.

**THEOREM 1.** *Suppose that  $\mathbf{A}$  is separable. If  $\mathbf{B}$  is a sufficient subfield, then there is a separable sufficient subfield  $\mathbf{B}_0$  satisfying*

$$\mathbf{B}_0 \subset \mathbf{B} \subset \mathbf{B}_0 \vee \mathbf{N}.$$

We recall that  $\mathbf{N}$  is the smallest  $\sigma$ -field containing the  $P$ -null sets. If the only  $P$ -null set is the empty set, then  $\mathbf{N} = \{\emptyset, X\} \subset \mathbf{B}_0$  and  $\mathbf{B}_0 \vee \mathbf{N} = \mathbf{B}_0$ . Accordingly, the following result is an immediate consequence of Theorem 1.

**COROLLARY 1.** *Suppose that  $\mathbf{A}$  is separable and the only  $P$ -null set is the empty set. If  $\mathbf{B}$  is sufficient, then  $\mathbf{B}$  is separable.*

**PROOF OF THEOREM 1.** Since  $\mathbf{A}$  is separable, there is a countable field  $\mathbf{A}_0$  such that  $\mathbf{A}$  is the smallest  $\sigma$ -field containing  $\mathbf{A}_0$ . Let  $\mathbf{B}$  be a sufficient subfield. Then, if  $A \in \mathbf{A}_0$ , there is a  $\mathbf{B}$ -measurable function  $g_A$  such that  $p(A \cap B) = \int_B g_A dp$ ,  $B \in \mathbf{B}$ ,  $p \in P$ . Let  $\mathbf{B}_0$  be the smallest  $\sigma$ -field with respect to which each of the functions  $g_A$ ,  $A \in \mathbf{A}_0$ , is measurable. Since  $\mathbf{A}_0$  is countable, it is clear that  $\mathbf{B}_0$  is separable. Also,  $\mathbf{B}_0 \subset \mathbf{B}$ .

Let  $\mathbf{A}_1$  be the collection such that  $A \in \mathbf{A}_1$  if and only if  $A \in \mathbf{A}$  and there is a  $\mathbf{B}_0$ -measurable function  $g$  satisfying

$$(4) \quad p(A \cap B) = \int_B g dp, \quad B \in \mathbf{B}, p \in P.$$

Then  $\mathbf{A}_0 \subset \mathbf{A}_1 \subset \mathbf{A}$ . Clearly,  $\mathbf{A}_1$  is a monotone class. Accordingly,  $\mathbf{A}_1 = \mathbf{A}$  since

$\mathbf{A}$  is the smallest monotone class containing  $\mathbf{A}_0$ . From the definition of  $\mathbf{A}_1$  and the relation  $\mathbf{B}_0 \subset \mathbf{B}$ , we conclude that  $\mathbf{B}_0$  is sufficient.

We now show that  $\mathbf{B} \subset \mathbf{B}_0 \vee \mathbf{N}$ . Suppose that  $A \in \mathbf{B}$ . Then  $A \in \mathbf{A} = \mathbf{A}_1$  and there is a  $\mathbf{B}_0$ -measurable function  $g$  satisfying (4). In particular,

$$0 = p(A \cap (X - A)) = \int_{X-A} g dp, \quad p \in P,$$

$$p(A) = p(A \cap A) = \int_A g dp, \quad p \in P.$$

Therefore, if  $h$  is the characteristic function of  $A$ , we have that  $g = h[P]$ , using the fact that  $0 \leq g \leq 1[P]$ , an immediate consequence of (4). Thus,  $h - g$  is  $\mathbf{N}$ -measurable, and  $h = g + (h - g)$ , being the sum of two  $\mathbf{B}_0 \vee \mathbf{N}$ -measurable functions, is  $\mathbf{B}_0 \vee \mathbf{N}$ -measurable. Consequently,  $A \in \mathbf{B}_0 \vee \mathbf{N}$ . This completes the proof.

In the following theorem let

$$a_x = \cap \{A \mid x \in A \in \mathbf{A}\}, \quad a_{0x} = \cap \{A \mid x \in A \in \mathbf{A}_0\},$$

where  $\mathbf{A}_0$  is a sufficient subfield. Let  $c$  be the cardinal number of the set of real numbers,  $c_0$  the cardinal number of the collection of sufficient subfields, and  $c_1$  the cardinal number of the collection of subfields containing  $\mathbf{A}_0$  that are not sufficient. Since we now know that  $0 < c_1$  is possible, it will not be too surprising to find out that sometimes  $c_0 < c_1$ .

**THEOREM 2.** *Suppose that  $\mathbf{A}$  is separable,  $\mathbf{A}_0$  is a sufficient subfield, the only  $P$ -null set is the empty set, and*

$$(5) \quad \text{card} \{a_{0x} \mid x \in X, a_x \neq a_{0x}\} \geq c.$$

*Then,*

$$c_0 \leq c < 2^c \leq c_1.$$

**PROOF.** By Corollary 1, each sufficient subfield must be separable. Therefore,

$$(6) \quad c_0 \leq \text{card} \{\mathbf{B} \mid \mathbf{B} \text{ is a separable subfield}\}.$$

Since  $\mathbf{A}$  is separable,  $\text{card } \mathbf{A} \leq c$  (see Problem 9 on page 26 of [5]). There is a one-to-one function from the set of separable subfields of  $\mathbf{A}$  to the set of countable subcollections of  $\mathbf{A}$ . If  $\mathbf{B}$  is a separable subfield, the value at  $\mathbf{B}$  of this function may be, for example, any particular countable subcollection of  $\mathbf{A}$  such that  $\mathbf{B}$  is the smallest  $\sigma$ -field containing the subcollection. Since  $\text{card } \mathbf{A} \leq c$ , the set of countable subcollections of  $\mathbf{A}$  has cardinal number less than or equal to  $c$ . Thus, the right hand side of (6) is less than or equal to  $c$ , implying that  $c_0 \leq c$ .

We now show that  $2^c \leq c_2$  where  $c_2$  is the cardinal number of the collection of subfields containing  $\mathbf{A}_0$ . Consequently,  $c_0 < c_2$ ,  $c_2 = c_2 - c_0$ ,  $c_1 = c_2$ , and  $2^c \leq c_1$ .

Let  $S$  be a subset of  $X$  such that  $X - S$  is the union of some subcollection of  $\{a_{0x} \mid x \in X, a_x \neq a_{0x}\}$ . Clearly, the collection of such sets  $S$  has cardinal number

greater than or equal to  $2^c$ , using (5). Let

$$(7) \quad \mathbf{B} = \{A \cup A_0 \mid A \subset S, A \in \mathbf{A}, A_0 \in \mathbf{A}_0\}.$$

Since  $\mathbf{A}$  is separable, if  $x \in X$  then  $a_x$  is the intersection of a countable number of sets in  $\mathbf{A}$  and hence is in  $\mathbf{A}$ . (Note that the partition  $\{a_x\}$  of  $X$ , induced by  $\mathbf{A}$ , is also the partition induced by any field with the property that  $\mathbf{A}$  is the smallest  $\sigma$ -field containing it. Here, since  $\mathbf{A}$  is separable, a countable field with this property exists.) Since  $\mathbf{A}_0$  is sufficient,  $\mathbf{A}_0$  is also separable. Accordingly,  $a_{0x} \in \mathbf{A}_0$ ,  $x \in X$ . From these facts, it follows that

$$(8) \quad S = \{x \mid a_x \in \mathbf{B}\}.$$

For if  $x \in S$  then  $a_x \subset a_{0x} \subset S$  and  $a_x$  is the union of a subset of  $S$  in  $\mathbf{A}$ , itself, with a set in  $\mathbf{A}_0$ , the empty set, and hence is in  $\mathbf{B}$ . If  $x$  is not in  $S$  then  $a_x \not\subset a_{0x} \subset X - S$ ,  $a_x$  is not in  $\mathbf{A}_0$  (for otherwise  $a_{0x} = a_x$ ), and  $a_x$  does not have the right form to be a set in  $\mathbf{B}$ . From (8) it follows that the mapping  $S \rightarrow \mathbf{B}$  described in (7) is one-to-one. Thus the cardinal number of the collection of  $\mathbf{B}$ 's is greater than or equal to  $2^c$ .

Let  $\mathbf{B}$  be as in (7). Then  $\mathbf{A}_0 \subset \mathbf{B} \subset \mathbf{A}$ . It remains to show that  $\mathbf{B}$  is a  $\sigma$ -field. Let  $B \in \mathbf{B}$ . Then there are sets  $A$  and  $A_0$  satisfying  $B = A \cup A_0$ ,  $A \subset S$ ,  $A \in \mathbf{A}$ ,  $A_0 \in \mathbf{A}_0$ . Let  $C_0 = \bigcup \{a_{0x} \mid x \in A\}$  and  $C = C_0 - A$ . If  $C_0$  is in  $\mathbf{A}_0$  then  $X - B$  is in  $\mathbf{B}$  and  $\mathbf{B}$  is a  $\sigma$ -field by the same reasoning as in Example 1. We now prove that  $C_0$  is in  $\mathbf{A}_0$ . Since  $\mathbf{A}_0$  is sufficient there is an  $\mathbf{A}_0$ -measurable function  $g$  satisfying

$$(9) \quad p(A \cap a_{0x}) = \int_{a_{0x}} g \, dp, \quad x \in X, p \in P.$$

If  $x \in X$  then  $g$  is constant on  $a_{0x}$  since  $g$  is  $\mathbf{A}_0$ -measurable and from (9) we have that

$$p(A \cap a_{0x}) = g(x)p(a_{0x}), \quad p \in P.$$

Since the only  $P$ -null set is the empty set, if  $A \cap a_{0x}$  is empty then  $g(x) = 0$  and if  $A \cap a_{0x}$  is nonempty then  $g(x) > 0$ . It is clear from the definition of  $C_0$  that if  $x \in C_0$  then  $A \cap a_{0x}$  is nonempty and if  $x$  is not in  $C_0$  then  $A \cap a_{0x}$  is empty. Thus,  $X - g^{-1}(\{0\}) = C_0$ , implying that  $C_0$  is in  $\mathbf{A}_0$ . This completes the proof.

REMARK 1. In Example 1,  $a_x = \{x\}$  and  $a_{0x} = \{x, -x\}$ , and it is clear that the conditions of Theorem 2 are satisfied. Many probability structures relevant for nonparametric statistical work satisfy the conditions, hence the conclusion, of Theorem 2. Among these, in addition to the one described in Example 1, the following is typical:

EXAMPLE 2. Let  $n$  be an integer  $> 1$ ,  $X$  Euclidean  $n$ -space,  $\mathbf{A}$  the collection of Borel subsets of  $X$ , and  $P$  the set of all probability measures  $p$  on  $\mathbf{A}$  of the form  $p = q \times \cdots \times q$ , where  $q$  is a probability measure on the  $\sigma$ -field of Borel subsets of the real line. If  $x = (x_1, \dots, x_n) \in X$ , let  $t_0(x)$  be the set of all points  $(x_{i_1}, \dots, x_{i_n})$ , where  $(i_1, \dots, i_n)$  is a permutation of  $(1, \dots, n)$ . Let  $\mathbf{A}_0$  be the subfield of  $\mathbf{A}$  induced by the statistic  $t_0$ . That is,  $\mathbf{A}_0$  is the collection

such that  $A \in \mathbf{A}_0$  if and only if  $A \in \mathbf{A}$  and there is a subset  $D$  of the range of  $t_0$  such that  $t_0^{-1}(D) = A$ . Here,  $a_{0x} = t_0(x)$  and  $a_x = \{x\}$ , and the assumptions of Theorem 2 are satisfied.

REMARK 2. With reference to Example 1, let  $t_0$  and  $t$  be functions on  $X$  satisfying  $t_0(x) = |x|$  if  $x \in X$ ,  $t(x) = x$  if  $x \in S$ ,  $= |x|$  if  $x \in X - S$ . The statistics  $t_0$  and  $t$  induce the subfields  $\mathbf{A}_0$  and  $\mathbf{B}$ , respectively, of Example 1. Since a statistic is sufficient if and only if its induced subfield is sufficient, we have that  $t_0$  is sufficient but  $t$  need not be sufficient. This is in spite of the fact that  $t_0 = F(t)$  for some function  $F$ .

Or with reference to Theorem 2, let  $t_0(x) = a_{0x}$  if  $x \in X$ ,  $t(x) = a_x$  if  $x \in S$ ,  $= a_{0x}$  if  $x \in X - S$ , where  $S$  is as described in the proof of Theorem 2. One can proceed as in the above paragraph and obtain a similar conclusion.

REMARK 3. Example 1, Theorem 2, and the above remarks indicate that sometimes a nonsufficient subfield or statistic can be as "informative" as a sufficient subfield or statistic. Accordingly, the definition of sufficiency in terms of conditional expectations, like most definitions, does not seem to capture all of the intuitive content commonly associated with the concept being defined. Needless to say, this, in itself, is not necessarily regrettable.

**3. Sufficiency in the general case.** Throughout this section, except in Example 3,  $(X, \mathbf{A}, P)$  is any probability structure. Making no further assumptions, we now prove several results about the sufficient subfields of  $\mathbf{A}$ . These results are easily shown to be true if  $P$  is assumed to be dominated. Without this assumption, these results and their proofs become somewhat more interesting.

THEOREM 3. Suppose that  $\mathbf{A}_1, \mathbf{A}_2, \dots$  are sufficient subfields.

(i) If  $\mathbf{A}_1 \supset \mathbf{A}_2 \supset \dots$ , then  $\bigcap_{n=1}^{\infty} \mathbf{A}_n$  is sufficient.

(ii) If  $\mathbf{A}_1 \subset \mathbf{A}_2 \subset \dots$ , then  $\bigvee_{n=1}^{\infty} \mathbf{A}_n$  is sufficient.

PROOF. Let  $f$  be a bounded  $\mathbf{A}$ -measurable function. There is, for each  $n$ , an  $\mathbf{A}_n$ -measurable function  $g_n$  such that  $g_n = E_p(f | \mathbf{A}_n)[p]$ ,  $p \in P$ . Let  $g(x) = \lim_{n \rightarrow \infty} g_n(x)$  for all  $x$  at which the limit exists,  $= 0$  otherwise.

Suppose that  $\mathbf{A}_1 \supset \mathbf{A}_2 \supset \dots$ . Then  $g$  is  $\bigcap_{n=1}^{\infty} \mathbf{A}_n$ -measurable. By the continuity theorem for conditional expectations [4, p. 331],  $\lim_{n \rightarrow \infty} g_n = E_p(f | \bigcap_{n=1}^{\infty} \mathbf{A}_n)[p]$ ,  $p \in P$ . Therefore,  $g = E_p(f | \bigcap_{n=1}^{\infty} \mathbf{A}_n)[p]$ ,  $p \in P$ . Hence,  $\bigcap_{n=1}^{\infty} \mathbf{A}_n$  is sufficient. The proof of (ii) is similar.

THEOREM 4. If  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are sufficient subfields and  $\mathbf{N}$  is contained in at least one of these subfields, then the subfield  $\mathbf{A}_1 \cap \mathbf{A}_2$  is sufficient.

PROOF. Suppose that  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are sufficient subfields and, without loss of generality, that  $\mathbf{N} \subset \mathbf{A}_2$ . If  $n$  is a positive integer let  $\mathbf{A}_{2n-1} = \mathbf{A}_1$  and  $\mathbf{A}_{2n} = \mathbf{A}_2$ . Let  $f$  be a bounded  $\mathbf{A}$ -measurable function. Define  $g_1, g_2, \dots$  inductively as follows: Let  $g_1$  be an  $\mathbf{A}_1$ -measurable function satisfying  $g_1 = E_p(f | \mathbf{A}_1)[p]$ ,  $p \in P$ . If  $g_{n-1}$  has been defined, let  $g_n$  be an  $\mathbf{A}_n$ -measurable function satisfying  $g_n = E_p(g_{n-1} | \mathbf{A}_n)[p]$ ,  $p \in P$ . Such a sequence  $g_1, g_2, \dots$  exists because  $\mathbf{A}_1, \mathbf{A}_2, \dots$  are sufficient subfields. Let  $g(x) = \lim_{n \rightarrow \infty} g_{2n-1}(x)$  for all  $x$  at which the limit exists,  $= 0$  otherwise. Let  $h(x) = \lim_{n \rightarrow \infty} g_{2n}(x)$  for all  $x$  at which the



limit exists,  $= 0$  otherwise. Then  $g$  is  $\mathbf{A}_1$ -measurable and  $h$  is  $\mathbf{A}_2$ -measurable. If  $p$  is in  $P$  let  $\mathbf{A}_{np}$  be the smallest  $\sigma$ -field containing  $\mathbf{A}_n$  and the collection of  $p$ -null sets. Then, by a theorem proved in [3],  $\lim_{n \rightarrow \infty} g_n = E_p(f | \mathbf{A}_{1p} \cap \mathbf{A}_{2p})[p]$ ,  $p \in P$ , implying that  $g = E_p(f | \mathbf{A}_{1p} \cap \mathbf{A}_{2p})[p]$ ,  $p \in P$ , and that  $\{x | g(x) \neq h(x)\}$  is in  $\mathbf{N}$ . Thus, since  $\mathbf{N} \subset \mathbf{A}_2$ , we have that  $g - h$  is  $\mathbf{A}_2$ -measurable. Therefore,  $g = h + (g - h)$  is the sum of two  $\mathbf{A}_2$ -measurable functions, hence is  $\mathbf{A}_2$ -measurable. Since  $g$  is measurable with respect to both  $\mathbf{A}_1$  and  $\mathbf{A}_2$ , it is  $\mathbf{A}_1 \cap \mathbf{A}_2$ -measurable. Moreover, for  $p \in P$ ,

$$\begin{aligned} g &= E_p(g | \mathbf{A}_1 \cap \mathbf{A}_2)[p] \\ &= E_p(E_p(f | \mathbf{A}_{1p} \cap \mathbf{A}_{2p}) | \mathbf{A}_1 \cap \mathbf{A}_2)[p] \\ &= E_p(f | \mathbf{A}_1 \cap \mathbf{A}_2)[p], \end{aligned}$$

since  $\mathbf{A}_1 \cap \mathbf{A}_2$  is contained in  $\mathbf{A}_{1p} \cap \mathbf{A}_{2p}$ . Thus,  $\mathbf{A}_1 \cap \mathbf{A}_2$  is sufficient.

**COROLLARY 2.** If  $\mathbf{A}_1, \mathbf{A}_2, \dots$  are sufficient subfields such that  $\mathbf{N} \subset \mathbf{A}_n$ ,  $n = 1, 2, \dots$ , then the subfield  $\bigcap_{n=1}^{\infty} \mathbf{A}_n$  is sufficient.

**PROOF.** Let  $\mathbf{B}_n = \bigcap_{k=1}^n \mathbf{A}_k$ . Then, by induction and Theorem 4, each subfield  $\mathbf{B}_n$  is sufficient. Applying part (i) of Theorem 3 now gives the desired result.

Consider the following two properties which a sufficient subfield  $\mathbf{A}_0$  may or may not have:

I. If  $\mathbf{B}$  is a sufficient subfield satisfying  $\mathbf{B} \subset \mathbf{A}_0$  then  $\mathbf{A}_0 \subset \mathbf{B}[P]$ .

II. If  $\mathbf{B}$  is a sufficient subfield then  $\mathbf{A}_0 \subset \mathbf{B}[P]$ .

A sufficient subfield  $\mathbf{A}_0$  satisfying (II) is sometimes termed a minimal sufficient subfield. It might be at least as appropriate, however, especially if the discussion is restricted to subfields containing  $\mathbf{N}$ , to use "least" or "smallest" in place of "minimal" and, instead, apply the adjective "minimal" to any sufficient subfield  $\mathbf{A}_0$  satisfying (I). Whether this is true or not hardly matters in the light of the following result:

**COROLLARY 3.** If  $\mathbf{A}_0$  is a sufficient subfield satisfying (I), then  $\mathbf{A}_0$  satisfies (II).

**PROOF.** Suppose that  $\mathbf{A}_0$  is sufficient and satisfies (I). Let  $\mathbf{B}$  be sufficient. Let  $\mathbf{A}_1 = \mathbf{B} \vee \mathbf{N}$ . It is easy to see that  $\mathbf{A}_1$  is sufficient. By Theorem 4,  $\mathbf{A}_0 \cap \mathbf{A}_1$  is sufficient, and, therefore, using (I),  $\mathbf{A}_0 \subset (\mathbf{A}_0 \cap \mathbf{A}_1) \vee \mathbf{N} \subset \mathbf{A}_1 \vee \mathbf{N} = \mathbf{B} \vee \mathbf{N}$ , the desired result.

**REMARK 4.** The condition involving  $\mathbf{N}$  in Theorem 4 cannot be eliminated entirely as the following example shows:

**EXAMPLE 3.** Let  $X$  be Euclidean 2-space,  $\mathbf{A}$  the collection of Borel subsets of  $X$ , and  $P$  the family of all probability measures  $p$  on  $\mathbf{A}$  satisfying  $p(D) = 1$  where

$$D = \{x | x = (x_1, x_2) \in X, x_1 = x_2\}.$$

For  $i = 1, 2$ , let  $\mathbf{A}_i$  be the subfield of  $\mathbf{A}$  induced by  $t_i$  where  $t_i(x) = x_i$ ,  $x \in X$ . It is easy to check that  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are sufficient but that  $\mathbf{A}_1 \cap \mathbf{A}_2 = \{\emptyset, X\}$  is not sufficient.

**REMARK 5.** The uncountable analogue of Corollary 2 is not true. That is,

there can exist a family  $H$  of sufficient subfields, each containing  $\mathbf{N}$ , such that  $\cap\{\mathbf{B} \mid \mathbf{B} \in H\}$  is not sufficient. Such an example is given by Pitcher [7]. If no such example existed, there would always exist a minimal sufficient subfield, contrary to fact [7].

LEMMA 1. *If  $\mathbf{B}$  is a sufficient subfield and  $A$  belongs to  $\mathbf{A}$ , then the smallest  $\sigma$ -field containing  $\mathbf{B} \cup \{A\}$  is sufficient.*

PROOF. Let  $\mathbf{C}$  be the smallest  $\sigma$ -field containing  $\mathbf{B} \cup \{A\}$ . Then

$$\mathbf{C} = \{(B_1 \cap A) \cup (B_2 \cap A') \mid B_i \in \mathbf{B}, i = 1, 2\}$$

where  $A' = X - A$ . For the purposes of this proof, if  $h$  is  $P$ -integrable let  $h'$  denote any  $\mathbf{B}$ -measurable function satisfying  $h' = E_p(h \mid \mathbf{B})[p]$ ,  $p \in P$ . Let  $r = 1 - s$  be the characteristic function of  $A$ .

We now show that  $\mathbf{C}$  is sufficient. Let  $f$  be an  $\mathbf{A}$ -measurable function into  $[0, 1]$ . Let

$$\begin{aligned} g_1(x) &= (rf)'(x)/r'(x) \quad \text{if } r'(x) \neq 0, \\ &= 0 \quad \text{if } r'(x) = 0, \end{aligned}$$

and let  $g_2$  be defined similarly using  $s$  in place of  $r$ . Then  $g_1$  and  $g_2$  are  $\mathbf{B}$ -measurable and

$$g = rg_1 + sg_2$$

is a  $\mathbf{C}$ -measurable function. Since  $0 \leq rf \leq r$ , we have that  $0 \leq (rf)' \leq r' [P]$ ,  $r'g_1 = (rf)' [P]$ , and  $0 \leq g_1 \leq 1 [P]$ . Similar results hold for  $s$  and  $g_2$ . Let  $C = (B_1 \cap A) \cup (B_2 \cap A')$  where  $B_i \in \mathbf{B}$ ,  $i = 1, 2$ . Then, for  $p \in P$ ,

$$\begin{aligned} \int_{B_1 \cap A} g \, dp &= \int_{B_1} rg_1 \, dp = \int_{B_1} (rf_1)' \, dp = \int_{B_1} r' g_1 \, dp \\ &= \int_{B_1} (rf)' \, dp = \int_{B_1} rf \, dp = \int_{B_1 \cap A} f \, dp. \end{aligned}$$

Similarly,  $\int_{B_2 \cap A'} g \, dp = \int_{B_2 \cap A'} f \, dp$ ,  $p \in P$ . Hence  $\int_C f \, dp = \int_C g \, dp$ ,  $p \in P$ .

The sufficiency of  $\mathbf{C}$  follows.

THEOREM 5. *If  $\mathbf{A}_1$  is a sufficient subfield and  $\mathbf{A}_2$  is a separable subfield, then  $\mathbf{A}_1 \vee \mathbf{A}_2$  is sufficient. In particular, if  $\mathbf{B}$  is a separable subfield containing a sufficient subfield, then  $\mathbf{B}$  is sufficient.*

PROOF. Let  $A_1, A_2, \dots$  be sets in  $\mathbf{A}$  such that  $\mathbf{A}_2$  is the smallest  $\sigma$ -field containing  $\{A_1, A_2, \dots\}$ . Let  $\mathbf{B}_0 = \mathbf{A}_1$  and define  $\mathbf{B}_1, \mathbf{B}_2, \dots$  inductively as follows: If  $n$  is a positive integer and  $\mathbf{B}_{n-1}$  has been defined, let  $\mathbf{B}_n$  be the smallest  $\sigma$ -field containing  $\mathbf{B}_{n-1} \cup \{A_n\}$ . Using Lemma 1, it follows that each of  $\mathbf{B}_1, \mathbf{B}_2, \dots$  is sufficient. Clearly,  $\mathbf{B}_1 \subset \mathbf{B}_2 \subset \dots$  and  $\bigvee_{n=1}^{\infty} \mathbf{B}_n = \mathbf{A}_1 \vee \mathbf{A}_2$ . Thus, by Theorem 3,  $\mathbf{A}_1 \vee \mathbf{A}_2$  is sufficient. The second assertion of the theorem is an immediate consequence of the first.

4. On the smallest subfield containing two sufficient subfields. Let  $\mathbf{A}_1$  and

$\mathbf{A}_2$  be sufficient subfields containing  $\mathbf{N}$ . Then  $\mathbf{A}_1 \cap \mathbf{A}_2$  is sufficient by Theorem 4. Is  $\mathbf{A}_1 \vee \mathbf{A}_2$  also sufficient? It is perhaps somewhat surprising to discover that  $\mathbf{A}_1 \vee \mathbf{A}_2$  need not be.

EXAMPLE 4. Let  $X$  be the set of all points  $x = (x_1, x_2)$  of Euclidean 2-space satisfying  $|x_1| = |x_2|$  and  $x_1 \neq 0$ . Let  $r_1(x) = (x_1, -x_2)$ ,  $r_2(x) = (-x_1, x_2)$ ,  $a_{ix} = \{x, r_i(x)\}$ ,  $x \in X$ , and  $\mathbf{A}_i$  be the smallest  $\sigma$ -field containing  $\{a_{ix} | x \in X\}$ ,  $i = 1, 2$ . Let  $\mathbf{B} = \mathbf{A}_1 \vee \mathbf{A}_2$ ,  $D = \{x | x \in X, x_1 = x_2\}$ , and  $\mathbf{A}$  be the smallest  $\sigma$ -field containing  $\mathbf{B} \cup \{D\}$ . If  $x \in X$ , let  $p_x$  be the probability measure on  $\mathbf{A}$  putting probability  $\frac{1}{2}$  on each of the points  $x, (x_1, -x_2), (-x_1, x_2), (-x_1, -x_2)$ . Finally, let  $P = \{p_x | x \in X\}$ . The set  $A$  is in  $\mathbf{A}_i$  if and only if there is a countable set  $S \subset X$  such that  $\bigcup \{a_{ix} | x \in S\}$  is either  $A$  or  $A'$ , primes being used to denote complements. A subset  $B$  of  $X$  is in  $\mathbf{B}$  if and only if  $B$  or  $B'$  is countable. Thus, if  $x \in X$  then  $\{x\} \in \mathbf{B}$  but  $D$  is not in  $\mathbf{B}$ . Clearly,

$$\mathbf{A} = \{(B_1 \cap D) \cup (B_2 \cap D') | B_i \in \mathbf{B}, i = 1, 2\}.$$

Here  $\mathbf{N} = \{\emptyset, X\}$ , hence is contained in any subfield.

Let  $i = 1$  or  $2$ . Then  $\mathbf{A}_i$  is sufficient, as we now show. Let  $f = f_1 + f_2$  where  $f_1$  is the characteristic function of  $B_1 \cap D$ ,  $f_2$  is the characteristic function of  $B_2 \cap D'$ , and  $B_1$  and  $B_2$  belong to  $\mathbf{B}$ . Let  $g_1 = f_1 + f_1(r_i)$ ,  $g_2 = f_2 + f_2(r_i)$ , and  $g = (g_1 + g_2)/2$ . If  $B_1$  is countable, then  $\{x | g_1(x) \neq 0\}$  is countable. If  $B_1'$  is countable, then  $\{x | g_1(x) \neq 1\}$  is countable. Therefore, in either case, since  $g_1 = g_1(r_i)$ ,  $g_1$  is  $\mathbf{A}_i$ -measurable. Similarly,  $g_2$  is  $\mathbf{A}_i$ -measurable implying that  $g$  is  $\mathbf{A}_i$ -measurable. If  $A_i \in \mathbf{A}_i$  and  $p \in P$ , it is clear that  $\int_{A_i} f(r_i) dp = \int_{A_i} f dp$ , implying that

$$\int_{A_i} g dp = \int_{A_i} \frac{1}{2}(f + f(r_i)) dp = \int_{A_i} f dp.$$

Therefore,  $\mathbf{A}_i$  is a sufficient subfield.

However,  $\mathbf{B} = \mathbf{A}_1 \vee \mathbf{A}_2$  is not sufficient. Otherwise, there would exist a  $\mathbf{B}$ -measurable function  $g$  satisfying  $p(D \cap B) = \int_B g dp$ ,  $B \in \mathbf{B}$ ,  $p \in P$ . In particular,  $p_x(D \cap \{x\}) = \int_{\{x\}} g dp_x$ ,  $x \in X$ , implying that  $g$  is the characteristic function of  $D$ . This is a contradiction since  $D$  is not in  $\mathbf{B}$ .

REMARK 6. The proof of Theorem 4 was based on a theorem proved in [3] which gives a simple way of obtaining the operator  $E_p(\cdot | \mathbf{A}_1 \cap \mathbf{A}_2)$  from the operators  $E_p(\cdot | \mathbf{A}_1)$  and  $E_p(\cdot | \mathbf{A}_2)$ . That there can be no closely analogous result for obtaining  $E_p(\cdot | \mathbf{A}_1 \vee \mathbf{A}_2)$  from  $E_p(\cdot | \mathbf{A}_1)$  and  $E_p(\cdot | \mathbf{A}_2)$  is implied by the above example.

Of course, certain extra assumptions, in addition to the assumption that  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are sufficient, imply that  $\mathbf{A}_1 \vee \mathbf{A}_2$  is sufficient. One such extra assumption is that  $P$  be a dominated family of measures. Another is that either  $\mathbf{A}_1$  or  $\mathbf{A}_2$  be separable (see Theorem 5). Still another is given in the following theorem.

THEOREM 6. Suppose that  $\mathbf{A}$  is separable. If  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are sufficient subfields, then  $\mathbf{A}_1 \vee \mathbf{A}_2$  is sufficient.

PROOF. By Theorem 1, there are separable sufficient subfields  $\mathbf{B}_1$  and  $\mathbf{B}_2$  such that

$$\mathbf{B}_i \subset \mathbf{A}_i \subset \mathbf{B}_i \vee \mathbf{N}, \quad i = 1, 2.$$

Therefore,  $B_1 \vee B_2$  is sufficient, by Theorem 5, and

$$B_1 \vee B_2 \subset A_1 \vee A_2 \subset B_1 \vee B_2 \vee N,$$

implying that  $A_1 \vee A_2$  is sufficient.

**COROLLARY 4.** *Suppose that  $A$  is separable. If  $A_1, A_2, \dots$  are sufficient subfields, then  $\bigvee_{n=1}^{\infty} A_n$  is sufficient.*

**PROOF.** It follows from Theorem 6 that  $\bigvee_{k=1}^n A_k$  is sufficient for each positive integer  $n$ . By Theorem 3, the desired result follows.

**5. Separability and sufficiency.** Separability of  $A$  or of one of its subfields plays an important role in Theorems 1, 5, 6, and elsewhere in the above sections. Even so, probably less can be said about sufficiency in the separable case than about sufficiency in the dominated case. Whether or not this is true, it should be kept in mind that nearly all, if not all, of the probability structures of importance in statistical work satisfy the condition that  $A$  is separable, but many do not satisfy the condition that  $P$  is dominated.

As usual, let  $(X, A, P)$  be any probability structure. Let  $D$  be the collection of Borel subsets of the real line. If  $B_0$  is a separable subfield of  $A$  then there is an  $A$ -measurable function  $f$  such that  $f^{-1}(D) = \{f^{-1}(D) \mid D \in D\} = B_0$ . (See Lemma 4 of [1], for example. Bahadur's blanket assumption that  $X$  is Euclidean, and so forth, is, of course, not needed and not used in his proof of Lemma 4.) Therefore, as an immediate consequence of Theorem 1, we have the following:

**THEOREM 7.** *Suppose that  $A$  is separable. If  $B$  is a sufficient subfield, then there is an  $A$ -measurable function  $f$  such that*

$$f^{-1}(D) = B[P].$$

This should be compared to a result of Bahadur: If  $A$  is separable,  $P$  is dominated, and  $B$  is a subfield, then there is an  $A$ -measurable function  $f$  such that  $f^{-1}(D) = B[P]$ . (This follows from Lemmas 3 and 4 of [1].) Theorem 7 indicates that if one adds the assumption that  $B$  is sufficient, then one can drop the assumption that  $P$  is dominated.

Of course, in Theorem 7 and the above,  $f$  could equally well be a measurable transformation into any Euclidean space with  $D$  again denoting the collection of Borel subsets of the space.

#### REFERENCES

- [1] R. R. BAHADUR, "Statistics and subfields," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 490-497.
- [2] R. R. BAHADUR, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423-462.
- [3] D. L. BURKHOLDER AND Y. S. CHOW, "Iterates of conditional expectation operators," *Proc. Amer. Math. Soc.*, Vol. 12 (1961), pp. 490-495.
- [4] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.
- [5] PAUL R. HALMOS, *Measure Theory*, D. Van Nostrand, New York, 1950.
- [6] PAUL R. HALMOS AND L. J. SAVAGE, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 225-241.
- [7] T. S. PITCHER, "Sets of measures not admitting necessary and sufficient statistics or subfields," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 267-268.

## ON A SPECIAL CLASS OF RECURRENT EVENTS

By M. P. SCHÜTZENBERGER

*Université de Poitiers and University of North Carolina*

**I. Introduction.** Let  $F$  be the set of all finite sequences (*words*) in the symbols  $x \in X$ . According to W. Feller ([2], Chap. VIII), a recurrent event  $\varepsilon$  is a pair  $(A, \mu)$  where  $A$  is a subset of  $F$  and  $\mu$  a probability measure fulfilling the conditions recalled below; one says that the event  $\varepsilon = (A, \mu)$  occurs at the last letter  $x_{i_n}$  of a word  $f = x_{i_1}x_{i_2} \cdots x_{i_n}$  if and only if  $f$  belongs to the set  $A$ ; we shall call  $A$  the *support* of  $\varepsilon$  and denote by  $T(A, \mu)$  the mean recurrence time of the event  $\varepsilon$ .

If the pair  $(B, \mu')$  defines another recurrent event on  $F$ , the pair  $(A \cap B, \mu')$  defines also a recurrent event. It results from the general theory of Feller ([2], Chap. VIII) that, when  $T(B, \mu')$  is finite, the ratio  $\pi = T(B, \mu')/T(A \cap B, \mu')$  is, in a certain sense, the limit of the conditional probability that a random word  $f \in F$  belongs to  $A$  when it is known to belong to  $B$ . For given arbitrary  $A$ , it is in general possible to find infinitely many  $(B, \mu')$  having finite  $T(B, \mu')$  which are such that  $\pi = 0$ .

The main point of this note is to verify several statements which, together, imply the following property:

**PROPERTY 1.** *If the support  $A$  is such that  $T(A \cap B, \mu')$  is finite for every recurrent event  $(B, \mu')$  having finite  $T(B, \mu')$ , then, for every such  $(B, \mu')$ ,  $\pi^{-1}$  is an integer at most equal to a certain finite number  $\delta^*$  which depends only upon  $A$ .*

Classical examples of this occurrence are the return to the origin in random walks over a finite group [3] and, in particular, the recurrent event which occurs at the end of every word whose length is an integral multiple of a particular integer.

In Section II, we discuss some properties of a class of recurrent events which we shall call *birecurrent*; in Section III, we verify the statements mentioned above, and in Section IV we describe examples of birecurrent supports.

**II. Preliminary remarks.** We consider  $F$  as the free monoid ([1], Chap. 1) generated by  $X$ ; the empty word  $e$  is the neutral element of  $F$  and the product  $ff'$  of the words  $f$  and  $f'$  is the word  $f''$  made up of  $f$  followed by  $f'$ ;  $f(f')$  is called a *left (right) factor* of  $f''$ ; a word is *proper* if it is different from  $e$ .

Feller's condition ([2], Chap. VIII) that the non empty subset  $A$  of  $F$  is the support of a recurrent event can be expressed as follows:  $U_r$ : if  $a \in A$  and  $f \in F$ , then,  $af \in A$  if and only if  $f \in A$ . This condition implies that  $A$  is a submonoid of  $F$  (i.e., that  $e \in A$  and  $A^2 \subset A$ ). We shall say that  $A$  is *birecurrent* if it satisfies  $U_r$  and the symmetric condition  $U_l$ ,  $U_l$ : if  $a \in A$  and  $f \in F$ , then,  $fa \in A$  if and only if  $f \in A$ .

It follows immediately that, if  $\{A_i\}$  is any collection of supports of recurrent

(birecurrent) events, the same is true of the intersection  $C$  of the sets  $A_i$ ; indeed,  $C$  is a submonoid because every  $A_i$  is a submonoid and, if, e.g.,  $a, af \in C$ , the word  $f$  belongs to all the sets  $A_i$  (because of  $U_r$ ) and consequently it belongs also to  $C$ .

Throughout this paper,  $A$  will denote a recurrent (or, eventually, birecurrent) support and we shall use the following notations:

$A^*$  = the set of all the proper words at the end of which the event whose support is  $A$  occurs for the first time; for any recurrent support  $B$ ,  $B^*$  is defined similarly.

$S = F - A^*F$  (= the complement in  $F$  of the right ideal  $A^*F$ );

$R = F - FA^*$ .

We state explicitly the following well known facts:

**II.1.** Every  $f \in F$  admits one and only one factorization  $f = as$  with  $a \in A$  and  $s \in S$  and at least one factorization  $f = ra'$  with  $a' \in A$  and  $r \in R$ . If and only if  $A$  is birecurrent the second factorization is unique for all  $f \in F$ .

**II.1'.** Every proper  $a$  of  $A$  admits a unique factorization as a product of elements of  $A^*$ .

The two statements are quite intuitive but a formal proof of them has been given in ([5]); II.1' shows that any bijection (i.e., one to one mapping onto) of  $A^*$  onto a set  $Y$  can be extended to an isomorphism of  $A$  onto the free monoid generated by  $Y$ .

The following remark will be used repeatedly in the course of this paper:

**II.1''.** When  $A$  is birecurrent, if  $s, s' \in S$  ( $r, r' \in R$ ) are such that  $s$  is a right factor of  $s'$  ( $r$  is a left factor of  $r'$ ) and that  $sf, s'f \in A$  ( $fr, fr' \in A$ ) for some  $f \in F$ , then  $s = s'$  ( $r = r'$ ). If, furthermore,  $f \in R$  ( $f \in S$ ), then  $sf \in A^* \cup \{e\}$ .

**PROOF.** Because of the perfect symmetry of  $U_r$  and  $U_l$  we can limit ourselves to the proof of the statement concerning  $s$  and  $s'$ . By hypothesis,  $s' = f's$  for some  $f' \in S$  and  $sf, f'sf \in A$ ; because of  $U_l$ , this implies  $f' \in A$ . Because of  $s' \in S = F - A^*F$  and II.1', this, in turn, implies  $f' = e$ , and we have proved that  $s' = es = s$ . Let us assume now that  $sr \in A$  with  $s \in S$  and  $r \in R$ . If, in addition,  $sr = e$ , the result is proved. If  $sr \in A - \{e\}$ , II.1' shows that  $sr = aa'$  with  $a \in A^*$  and  $a' \in A$ ; as above,  $a$  cannot be a left factor of  $s$  and, consequently,  $a'$  is a right factor of  $r$ ; but, by a symmetrical argument, this shows that  $a' = e$  and that consequently  $sr = a \in A^*$ . This concludes the proof of II.1''.

Let us assume now that  $A$  is birecurrent; we denote by  $\Delta Sf(\Delta Rf)$  the set of the right (left) factors of  $f$  that belong to  $S(R)$  and by  $\Delta f$  the set of the triples  $(r, a, s)$  such that  $f = ras$  and that  $r \in R, a \in A, s \in S$ ; such a triple will be called an  $A$ -factorization of  $f$  and  $\delta f$  will denote the number of distinct triples in the set of the  $A$ -factorizations of  $f$ .

**II.2.** For any  $f, f' \in F$ ,  $\delta ff' \geq \max(\delta f, \delta f')$  and  $\delta ff' = \delta f (= \delta f')$  if and only if for every left (right) factor  $f''$  of  $f'$  (of  $f$ ) the product  $ff''$  ( $f''f'$ ) has a factorization  $ff'' = sa$  ( $f''f' = ar'$ ) where  $a \in A$  and where  $f''$  is a right (left) factor of  $a$ .



PROOF. Let us consider any element  $g \in F$  and prove that there exists a bijection  $\sigma_g: \Delta Rg \rightarrow \Delta Sg$ . Indeed, by II.1, to any  $r \in \Delta Rg$  (i.e., to any  $r \in R$  which is such that  $g = rg'$  for some  $g' \in F$ ) there corresponds a unique  $s \in \Delta Sg$  (determined by the conditions  $g' = as$ ,  $a \in A$ ,  $s \in S$ ) which we call  $\sigma_g r$ ; because of the symmetry implied by the hypothesis that  $A$  is birecurrent we can construct in a similar manner a mapping  $\Delta Sg \rightarrow \Delta Rg$  which we call  $\sigma_g^{-1}$ . Since, clearly, for any  $r \in \Delta Rg$  we have  $(\sigma_g^{-1} \circ \sigma_g)r = r$  and similarly for any  $s \in \Delta Sg$ , this shows that  $\sigma_g$  is a bijection and also that the  $A$ -factorizations of  $g$  are in a one-to-one correspondence with the elements of  $\Delta Rg$ .

We now revert to the proof of II.2. By the above construction we know that  $\delta f f'$  is equal to  $\delta f$  (i.e., to the number of elements in  $\Delta Rf$ ) plus the number of proper  $r' \in \Delta Rf'$  such that  $fr' \in R$ . Thus,  $\delta f f' \geq \delta f$  with the equality sign if and only if we do not have  $f f'' \in R - \Delta Rf$  for some left factor  $f''$  of  $f'$ , i.e., if and only if every such  $f f''$  satisfies the condition stated in II.2. Because of the symmetry this concludes the proof.

For any  $f \in F$ , let us denote by  $\alpha f$  the smallest positive integer for which  $f^{\alpha f} \in A$ ;  $\alpha f$  is infinite if the only finite power of  $f$  that belongs to  $A$  is  $f^0 (= e$ , by definition).

II.3. A sufficient condition that the recurrent support  $A$  is birecurrent is that  $\alpha f$  is finite for all  $f \in F$ ; reciprocally if  $A$  is a birecurrent support, then, for any  $f \in F$ ,  $\alpha f$  is at most equal to the supremum  $\delta' f$  of  $\delta f^m$  over all the positive powers of  $f$ .

PROOF. By hypothesis,  $A$  satisfies  $U_r$  and, in order to show that it is birecurrent, it will be enough to show that if  $a$  and  $fa$  belong to  $A$  then  $f$  also belongs to  $A$ . Let us assume that  $(af)^m \in A$  for some positive finite  $m$ ; we have  $(af)^m = a(fa)^{m-1}f \in A$  and, because of the fact that  $a$ ,  $(fa)^{m-1} \in A$  and  $U_r$ , this implies  $f \in A$ . This proves the first part of II.3.

Now let  $A$  be birecurrent and  $f$  such that  $\delta' f$  is finite; by II.1, any  $f^n$  ( $0 \leq n \leq \delta' f$ ) admits an  $A$ -factorization  $(e, a_n, s_n)$  and, by II.2, to each such  $s_n$  there corresponds one  $A$ -factorization of  $f^{s' f}$ . Since, by definition,  $\delta f^{s' f} \leq \delta' f$ , we must have  $s_n = s_m (= s$ , say) with  $0 \leq m, n \leq \delta' f$  and, e.g.,  $m < n$ . Thus,  $f^n = as$  and  $f^m = a's$  with  $a, a' \in A$  and, after cancelling  $s$ , we obtain  $f^{n-m}a' = a$ . Because of  $U_l$ , this last relation shows that  $f^{n-m}$  belongs to  $A$  and, since  $0 < n - m \leq \delta' f$ , by construction, the result is entirely proved.

Let us assume now that  $A$  is birecurrent and that  $f$  is such that  $\delta f = \delta f^2 < \infty$ . We consider the set  $K$  (containing at least  $f^2$ ) defined by  $K = \{f' \in fFf: \delta f' = \delta f\}$ .

II.4. There exists a group  $G$ , a subgroup  $H$  of  $G$  and a mapping  $\sigma: K \rightarrow G$  that have the following properties:  $\sigma$  is an epimorphism (i.e., homomorphism onto) and  $G$  is finite;  $\sigma^{-1}H = K \cap A$  and the index of  $H$  in  $G$  is at most  $\delta f$ .

PROOF. According to II.2, the hypothesis  $\delta f = \delta f^2$  implies the existence of a bijection  $\sigma^*: \Delta Sf \rightarrow \Delta Rf$  defined for each  $s \in \Delta Sf$  by  $\sigma^*s$ , the unique  $r \in \Delta Rf$  which is such that  $sr \in A$ ; trivially,  $\sigma^*e = e$ . Also, by II.2 and the very definition of  $K$ , we have  $\Delta Rk = \Delta Rf$  and  $\Delta Sk = \Delta Sf$  for any  $k \in K$ ; consequently,  $K^2 \subset K$ .

Thus, recalling the definition of  $\sigma_f$  given in the proof of II.2, we can associate to any  $k \in K$  a bijection  $\sigma_k^*: \Delta Rf = \Delta Rf$  defined by  $\sigma_k^* = \sigma^* \circ \sigma_k$ .

Let us now verify that for any  $k, k' \in K$  we have  $\sigma_{kk'}^* = \sigma_k^* \circ \sigma_{k'}^*$ . Indeed, if  $(r, a, s) \in \Delta k$  and  $(r', a', s') \in \Delta k'$  we shall have  $(r, a'', s') \in \Delta kk'$  for some  $a'' \in A$  if and only if  $sr' \in A$  and the identity is verified. Because of the hypothesis that  $\delta f$  is finite, this construction shows that the set  $\{\sigma_k^*\} (k \in K)$  is a group  $G$  and that the mapping  $\sigma$  which sends every  $k \in K$  onto  $\sigma_k^*$  is an epimorphism.

Observe now that  $k$  belongs to  $A$  if and only if  $(e, k, e) \in \Delta k$ , that is, if and only if  $\sigma_k^*$  keeps  $e$  invariant. Again, because  $G$  is finite, the elements  $k \in K$  which have this last property map onto a subgroup  $H$  of  $G$  and, clearly,  $\sigma^{-1}H$  is contained in  $A$ . The fact that the index of  $H$  in  $G$  is at most equal to the number of elements in  $\Delta Rf$  (i.e., to the number  $\delta f$ ) is a standard result from group theory. As a corollary of II.4 we state II.4'.

**II.4'.** If  $A$  is such that the supremum  $\delta^*$  of  $\delta f'$  over all  $f' \in F$  is finite and if  $\delta f = \delta^*$ , then the representation  $\{\sigma_k^*\}$  described in II.4 is isomorphic to the representation of  $G$  over the cosets of  $H$ .

**PROOF.** The property stated amounts to the statement that the group  $G = \{\sigma_k^*\}$  is transitive or, in an equivalent fashion, to the fact that for every  $s \in \Delta Sf$  there exists at least one  $k \in K$  such that  $\sigma_k e = s$ , i.e., such that  $k = as$  with  $a \in A$ .

In order to prove this, let  $(r, a', s) \in \Delta f$ . By II.3 we know that there exist finite positive integers  $m$  and  $m'$  such that  $f^m \in A$  and  $r^{m'} \in A$ . Thus the product  $f^m f^{m'-1} f = f^m f^{m'} as$  admits the factorization  $a''s$  with  $a'' = f^m f^{m'} a' \in A$  and it belongs to  $K$  since, under the hypothesis that  $\delta f$  is maximal,  $K$  is identical to  $fFf$ .

The next statement is not needed for the verification of property 1. Its aim is to show that the representation described in Section IV below covers all the birecurrent supports with finite  $\delta^* = \sup \delta f$ .

**II.5.** If  $A$  is a birecurrent support with finite  $\delta^*$  there exists a monoid  $M$  and an epimorphism (homomorphism onto)  $\gamma: F \rightarrow M$  such that  $\gamma^{-1}\gamma A = A$ , and that  $M$  admits minimal ideals.

**PROOF.** Let us consider any  $f \in F$  and denote by  $\{\gamma f\}$  the set of all  $f' \in F$  which satisfy the following condition: for any  $f_1, f_2 \in F$ ,  $f_1 f f_2 \in A$  if and only if  $f_1 f' f_2 \in A$ . The relation  $f' \in \{\gamma f\}$  is reflexive and transitive and it is well known that it is compatible with the multiplicative structure of  $F$  (i.e., it is a congruence); thus we can identify each set  $\{\gamma f\}$  with an element  $\gamma f$  of a certain quotient monoid  $M$  of  $F$ . Since  $f \in A$  if and only if  $f_1 f f_2 \in A$  with  $f_1 = f_2 = e$ ,  $A$  is the union of the sets  $\{\gamma a\} (a \in A)$  and, trivially,  $\gamma^{-1}\gamma A = A$ .

Let us now take an element  $f$  such that  $\delta f = \delta^*$ , a finite quantity; according to II.2, the maximal character of  $\delta f$  implies that for every  $f_1$  the product  $f_1 f$  has a left factor  $f_1 r \in A$  for some  $r \in \Delta Rf$ . Thus, because of the symmetry, any relation  $f_1 f f_2 \in A$  implies  $f_1 r, f_2 \in A$  with  $(r, a, s) \in \Delta f$ .

It follows immediately that for any two  $k, k' \in K (= fFf)$ , the relation  $\gamma k = \gamma k'$  is equivalent to the relation  $\sigma k = \sigma k'$  in the notations of II.4. Thus,  $\sigma K$  is

isomorphic to a group and since  $K$  is the intersection of a right and of a left ideal of  $F$ , this shows that  $M$  admits minimal ideals.

We now revert to the preparation of the proof of the main property and we consider  $A$ , a birecurrent support,  $B$  a recurrent support and  $C = A \cap B$ ; we assume that  $C$  does not reduce to  $\{e\}$  and that consequently  $C^*$  (the set of the proper words at the end of which the events whose supports are  $A$  and  $B$  respectively occur together for the first time) is not empty.

**II.6.** Any element  $f$  from  $F - C^*F$  has a unique factorization  $f = f_1 f_2$  with  $f_1 \in B - C^*B$  and  $f_2 \in F - B^*F$ ; conversely any such product  $f_1 f_2$  belongs to  $F - C^*F$ .

**PROOF.** Because of II.1 any  $f$  has a unique factorization  $f = f_1 f_2$  with  $f_1 \in B$  and  $f_2 \in F - B^*F$ . Since  $C$  is a recurrent support contained in  $B$ , any product  $f_1 f_2$  with  $f_1 \in B$  and  $f_2 \in F - B^*F$  belongs to  $F - C^*F$  if and only if  $f_1$  belongs to  $B - C^*B$  and this concludes the proof.

As mentioned in II.1', there exists an isomorphism  $\beta: B \rightarrow Q$  where  $Q$  is the free monoid generated by  $Q^* = \beta B^*$  and it is easily verified that the image  $P$  of  $C$  by  $\beta$  satisfies  $U_r$  and  $U_l$  when, according to our hypothesis,  $A$  is birecurrent. Indeed,  $P$  is surely a submonoid of  $Q$  and it is enough to verify that the relations  $p, p', pqp' \in P$  imply  $q \in Q$  (because  $\beta^{-1}p, \beta^{-1}p', \beta^{-1}pqp' \in A$  imply, e.g.,  $\beta^{-1}qp \in A$ , by  $U_r$ , then  $\beta^{-1}q \in A$ , by  $U_l$  and, finally  $q \in P = \beta(A \cap B)$ ).

As before, we define a  $P$ -factorization of an element  $q \in Q$  as a triple  $(\bar{r}, p, \bar{s})$  such that  $q = \bar{r}p\bar{s}$  and that  $\bar{r} \in \bar{R} = Q - QP^*$ ,  $p \in P$ ,  $\bar{s} \in \bar{S} = Q - P^*Q$  with  $P^* = \beta C^*$ . All the remarks made in II.2 apply here since  $P$  is a birecurrent support in  $Q$ , and we define  $\bar{\delta}q$  as the number of  $P$ -factorizations of  $q$ .

**II.7.** For any  $b \in B$ ,  $\bar{\delta}pb \leq \bar{\delta}b$ .

**PROOF.** Let  $\bar{r}$  be any element of  $\bar{R}$  and define  $\beta^*\bar{r}$  as the (uniquely determined) element  $r \in R$  such that  $(r, a, e) \in \Delta b$  for some  $a \in A$ . We show that the restriction of the mapping  $\beta^*$  to any set  $\Delta\bar{R}q$  ( $q \in Q$ ) is an injection (i.e., is one to one into). Indeed, if  $\bar{r}, \bar{r}' \in \Delta\bar{R}q$  we have, e.g.,  $\bar{r}' = \bar{r}q'$  for some  $q' \in Q$ ; thus, if  $\beta^*\bar{r} = \beta^*\bar{r}'$  ( $= r$ , say), we have the following relations:  $\beta^{-1}\bar{r} = ra \in B$  with  $a \in A$ ;  $\beta^{-1}\bar{r}' = ra' \in B$  with  $a' \in A$ ;  $ra' = rab'$  with  $b' = \beta^{-1}q\beta \in B$ . Consequently,  $a' = ab'$  and, because of  $U_r$ ,  $b' \in A$ . This shows that  $q' = \beta b'$  belongs to  $P$  and that finally,  $q' = e$  because of the relation  $\bar{r}' = \bar{r}q' \in \bar{R}$ . Thus,  $\bar{r}' = \bar{r}$  and our contention is proved.

The remark II.7 is also proved since we have shown that for any  $b \in B$  there exists an injection of  $\Delta\bar{R}\beta b$  into  $\Delta Rb$ .

**II.8.** If  $\delta^* (= \sup \delta f)$  is finite and if  $\delta b = \delta^*$  for at least one  $b \in B$ , then  $\bar{\delta}^* (= \sup \bar{\delta}q)$  is a divisor of  $\delta^*$ .

**PROOF.** Under these hypotheses, we may assume without loss of generality that  $B$  contains an element  $f$  such that  $\delta f = \delta^*$  and  $\bar{\delta}\beta f = \bar{\delta}^*$ . We use the notations of II.4 and II.4'. By construction, the image  $G'$  by  $\sigma$  of  $B \cap K$  is a subgroup of  $G$  and we have  $B \cap \sigma^{-1}(H \cap G') = A \cap B \cap K$ . Thus, by a standard result of

group theory the index  $\delta'^*$  of  $H \cap G'$  in  $G'$  is a divisor of the index of  $H$  in  $G$  (i.e., of  $\delta^*$ ). We prove now that  $\delta'^*$  is in fact equal to  $\delta^*$ ; for this we repeat the construction of II.4 and II.4' with  $\beta(B \cap K)$  in the role of  $K$  and we obtain an epimorphism  $\bar{\sigma}: \beta(B \cap K) \rightarrow \bar{G}$  such that  $\bar{\sigma}^*$  is the index of the subgroup  $\bar{H}$  of  $\bar{G}$ . We recall the definition of the mapping  $\beta^*$  used in II.7 and we observe that we can define a bijection  $\beta^{*-1}: \Delta Rf \cap \beta^* \Delta \bar{R} \beta f \rightarrow \Delta \bar{R} \beta f$  such that  $\beta^{*-1} \circ \beta^*$  is the identity mapping of  $\Delta \bar{R} \beta f$  onto itself;  $\beta^{*-1}$  induces in a natural fashion an epimorphism  $\beta^{**}: G' \rightarrow \bar{G}$  and, trivially,  $H \cap G'$  is the inverse image of  $\bar{H}$  by  $\beta^{**}$ . Thus  $\bar{\sigma}^*$  is equal to  $\delta'^*$  and II.8 is proved.

**III. Verification of property 1.** We keep the notations already introduced and we assume that  $(A, \mu)$  is a recurrent event. According to Feller,  $\mu$  satisfies the two conditions:

$$M_0: \mu e = 1 \text{ and for any } f \in F, \mu f = \sum (\mu f x: x \in X),$$

$$M_r: \text{if } a \in A \text{ and } f \in F \text{ then } \mu af = \mu a \mu f.$$

We shall say that  $\mu$  is a *positive product measure* if  $\mu ff' = \mu f \mu f' > 0$  for any  $f, f' \in F$ , and, in this case,  $M_r$  is trivially satisfied.

We denote by  $|f|$  the length of the element  $f$  and for any subset  $F'$  of  $F$  we use the following notations:  $F'_n = \{f \in F': |f| \leq n\}$ ;  $\mu F' = \lim_{n \rightarrow \infty} \sum \{\mu f: f \in F'_n\}$ . It follows that  $\mu F' \leq 1$  if  $F'$  is such that any  $f \in F$  has at most one left factor which belongs to  $F'$ ; this condition is satisfied in particular by any subset of  $A^*$  and, according to Feller's definition, we shall say that  $(A, \mu)$  is *persistent* if and only if  $\mu A^* = 1$ . The next two statements are verified by an imitation of Feller's proof procedure.

**III.1.** For any recurrent event  $(A, \mu)$  we have  $T(A, \mu) = \mu S$ .

**PROOF.** Let us introduce for any  $s \in S$  the notation  $S(s) = S \cap sF$ . We verify the identities

$$(III.1). \text{ for all } m \geq |s|: 0 \leq \mu s - \mu A_{m+1}^*(s) = \mu S_{m+1}(s) - \mu S_m(s);$$

$$(III.1'). \text{ for all } m \geq 1: (1 - \mu A^*) + (\mu A^* - \mu A_m^*) = \mu S_m - \mu S_{m-1}$$

Indeed, (III.1) is an immediate consequence of  $M_0$  and of the fact that the sets  $\{s\} \cup S_m(s)X$  and  $S_{m+1}(s) \cup A_{m+1}^*(s)$  are identical for any  $m \geq |s|$ . (III.1') is the special case of (III.1) for  $s = e$ .

From this second identity we deduce that if  $\mu A^* = 1$  we have  $\lim_{m \rightarrow \infty} (\mu S_m - \mu S_{m-1}) = 0$ . Thus, *a fortiori* (from the first identity)  $\mu A^* = 1$  implies  $\mu s = \mu A^*(s)$ . We now sum the second identity from  $m = 1$  to  $m = n$ . After rearranging terms, we obtain:

$$(III.1''). \quad \mu S_n = (n+1)(1 - \mu A_n^*) + \sum \{|a| \mu a: a \in A_n^*\}.$$

This shows that if  $(A, \mu)$  is not persistent,  $\mu S$  is infinite and we assume now that  $\mu A^* = 1$ . Under this hypothesis,  $T(A, \mu)$  is defined as  $\lim_{n \rightarrow \infty} \sum \{|a| \mu a: a \in A_n^*\}$ , and since  $\mu A^* = 1$  implies that

$$(n+1)(1 - \mu A_n^*) = \sum \{(n+1) \mu a: a \in A^* - A_n^*\},$$

we can write for all  $n$

$$\sum \{|a| \mu a: a \in A_n^*\} \leq \mu S_n \leq \sum \{|a| \mu a: a \in A^* - A_n^*\} + \sum \{|a| \mu a: a \in A_n^*\}.$$

This concludes the proof since it shows that  $\mu S = T(A, \mu)$  if this last quantity is finite and that  $\mu S$  is infinite if  $T(A, \mu)$  is so.

For any  $s \in S$  let us define  $R^*(s)$  as  $\{e\}$  if  $s = e$  and, as the set of those  $f \in F$  such that  $sf \in A^*$ , if  $s \neq e$ .

**III.2.** If  $A$  is birecurrent,  $\mu$  a product measure and  $(A, \mu)$  persistent, we have  $T(A, \mu) = \mu R$  and, for all  $s \in S$ ,  $1 = \mu R^*(s)$ .

**PROOF.** Under these hypotheses all the notions are perfectly symmetrical. Thus, the identity (III.1'') shows that  $\mu R_n = \mu S_n$  and, as a special case, that  $\mu R = T(A, \mu)$ . Since any  $a \in A^*(s)$  has a unique factorization  $a = sf$  with  $f \in R^*(s)$ , and since  $\mu$  is a product measure, we have for all  $m \geq |s|$  the identity

$$(III.2) \quad \mu A_m^*(s) = \mu s \mu R_{m-|s|}^*(s).$$

Thus, we have in any case  $\mu R(s) = \mu A^*(s)/\mu s \leq 1$  because of the formula (III.1); with the equality sign when  $(A, \mu)$  is persistent because as seen above  $\mu s = \mu A^*(s)$ .

**III.3.** If  $A$  is birecurrent and  $\mu$  a product measure,  $T(A, \mu) = \delta^*$ .

**PROOF.** We use the notations of Section II and we recall the following facts:

(1) According to II.1'',  $R^*(s)$  is a subset of  $R$ ;

(2) for the same reason, if  $s, s' \in \Delta Sf$  for some  $f \in F$ , the sets  $R^*(s)$  and  $R^*(s')$  are disjoint.

(3) if  $\delta^*$  is finite and  $\delta f = \delta^*$  then, by II.2, to every  $r \in R$  there corresponds one  $s \in \Delta Sf$  such that  $sr \in A^*$ . Thus, in this case, the union of the sets  $R^*(s)$  over all  $s \in \Delta Sf$  is equal to  $R$ . Now to the proof! We shall show that if  $\delta f = \delta^*$  we have the inequalities  $\mu R \leq \delta f \leq \mu R$  and, trivially, the result will follow by III.2.

The second inequality is vacuously true when  $(A, \mu)$  is not persistent since, then,  $\mu R$  is infinite. When  $(A, \mu)$  is persistent we have for any  $f' \in F$  the inequality  $\delta f' = \sum \{\mu R^*(s): s \in \Delta Sf'\} \leq \mu R$  since, then,  $\mu R^*(s) = 1$  and since the sets  $R^*(s)$  are pairwise disjoint. Thus the second inequality is always true. If now  $\delta f = \delta^*$ , we know by 3 above that  $\sum \{\mu R^*(s): s \in \Delta Sf\} = \mu R$ . Since in any case, as we have seen in the proof of III.2, we have  $\mu R^*(s) \leq 1$ , it follows that  $\mu R \leq \delta^*$  and the result is proved.

**III.4.** If  $(B', \mu)$  is a recurrent event and if  $A$  is birecurrent we have

$$T(A \cap B', \mu) = \bar{\delta}^* T(B', \mu)$$

where  $\bar{\delta}^*$  is defined below.

**PROOF.** Let  $B = \{b \in B': \mu b > 0\}$  and  $C = A \cap B$ ; it is easily verified that  $(B, \mu)$  is again a recurrent event and that, according to III.1, we have

$$\begin{aligned} T(A \cap B', \mu) &= T(A \cap B, \mu) = \mu(F - C^*F) \\ T(B', \mu) &= T(B, \mu) = \mu(F - B^*F). \end{aligned}$$



We keep the notations used in the proofs of II.6 and II.7 and we observe that, by taking into account II.6 and the condition  $M_r$  on  $\mu$ , the remark III.4 is equivalent to the relation  $\mu(B - C^*B) = \bar{\delta}^*$ . In order to prove this identity we define a measure  $\nu$  on  $Q$  by the relation  $\nu\beta b = \mu b$ , for all  $b \in B$ ; because of  $M_r$  and of the definition of  $B$ ,  $\nu$  is a positive product measure and, since we know that  $P = \beta C$  is birecurrent,  $(P, \nu)$  is a recurrent event on  $Q$ . Because of III.1 and III.3  $T(P, \nu) = \nu(Q - P^*Q) = \bar{\delta}^*$ . But, by definition,  $\nu(Q - P^*Q) = \nu\beta(B - C^*B) = \mu(B - C^*B)$  and the result is proved.

**III.5.** If  $\delta^*$  is finite, and  $(B', \mu)$  persistent for some measure  $\mu$  which satisfies the condition that for every  $f \in F$  at least one element from  $FfF$  has positive measure, then  $\bar{\delta}^*$  is a divisor of  $\delta^*$ .

**PROOF.** Because of the conditions satisfied by  $\mu$  and  $\delta^*$  we can find an element  $f$  such that  $\delta f = \delta^*$  and that  $\mu f > 0$ ; we have  $f = b's'$  with  $b' \in B$  and  $s' \in F - B^*F$ . Because  $(B, \mu)$  is persistent, it follows from III.1 that  $\mu(B^* - s'F) = \mu s'$ . Since this last quantity is positive, there exists at least one element  $b \in B^* \cap s'F$ . Finally, because of II.2 we have  $\delta b'b = \delta^*$  with  $b'b \in B$ . Thus, we can apply II.8 and the result is proved.

The next statement is intended to give a characterization of the birecurrent supports in terms of their intersection with other recurrent events; by  $E$  we mean any fixed birecurrent support such that  $T(E, \mu)$  is finite for some positive product measure  $\mu$ ;  $E^*$  is defined as usual and we say that  $(E', \mu')$  belongs to the family  $((E))$  if the two following conditions are met:

- (i).  $(E', \mu')$  is a recurrent event on  $F$ ;
- (ii). there exists a finite integer  $m$  such that any element from  $E'^*$  is the product of  $m$  words from  $E^*$ . It is trivial that under these hypotheses  $E'$  is birecurrent. Since  $F$  itself is a birecurrent support (with  $F^* = X$ ) a simple example of a family  $((E))$  is the family of the birecurrent events  $(F_{(m)}, \mu_m)$  where  $F_{(m)}$  is the set of all words whose length is a multiple of  $m$  and where  $\mu_m$  is a suitable measure.

**III.6.** If the recurrent support  $A$  is such that  $(A \cap E', \mu')$  is persistent for every  $(E', \mu') \in ((E))$ , then,  $A$  is a birecurrent support.

**PROOF.** This is a simple application of II.3 and we use the notations of this remark. If  $\alpha f$  is finite for all  $f$ , then we know by II.3 that  $A$  is birecurrent. Thus we may suppose that  $A$  and  $f$  are such that  $\alpha f$  is infinite and we show that  $(A \cap E', \mu')$  is not persistent for some suitable  $(E', \mu')$ . Indeed, by the second part of II.3 we know that  $f^m \in E$  for some finite positive  $m$ . Thus  $f^m$  admits a factorization as a product of  $m'$  elements from  $E^*$ . We take  $E'$  defined by the condition  $E'^* = E^{*m'}$  and  $\mu'$  defined by the condition that  $\mu' f^m = 1$  and  $\mu' f' = 0$  for any other  $f' \in E'^*$ . The conditions  $M_0$  and  $M_r$  recalled at the beginning of this section are obviously satisfied and  $T(E', \mu')$  is finite. Finally,  $(A \cap E', \mu')$  cannot be persistent since  $A \cap E'$  reduces to  $\{e\}$  and this ends the proof.

Clearly, the conditions of III.6 are satisfied if  $A$  is such that  $T(A \cap B, \mu) < \infty$  for any  $(B, \mu)$  with finite  $T(B, \mu)$ .

The next statement is a simple application of II.2.



**III.7.** If  $A$  is birecurrent and if  $\delta^*$  is finite, then, for any product measure,  $\mu$ , the distribution of the recurrence time of  $(A, \mu)$  has moments of every order.

**PROOF.** Let  $A' = \{a \in A : \mu a > 0\}$ . Trivially,  $A'$  is birecurrent and, by II.7 we know that every  $f \in F$  has at most  $\delta^* A'$ -factorizations. Since the distribution of the recurrence times of  $(A, \mu)$  and  $(A', \mu)$  are the same, there is no loss of generality in assuming that  $A = A'$ , i.e., that  $\mu$  is positive.

Since  $\delta^*$  is finite there exists an element  $f \in F$  which, because of II.2, has the property that for any proper  $s \in S$  the product  $sf$  has a factorization  $sf = ar$  with  $a \in A^*A$ . Thus, for any integer  $n$ , the definition  $S = F - A^*F$  allows us to write the inequality

$$\Sigma\{\mu f' : f' \in A^*, n|f| < |f'| \leq (n+1)|f|\} = \mu A_{(n+1)}^*|f| - \mu A_n^*|f| \leq (1 - \mu f)^{n+1}.$$

Consequently the distribution of the  $|a|$  for  $a \in A^*$ , i.e., of the recurrence time of  $A^*$ , is dominated by an exponential distribution and this proves the result.

**IV. Examples.** We want to describe a class of monoids,  $V$ , which allows the construction of birecurrent supports. For this purpose, we consider a group  $G'$  (whose elements are identified with the corresponding elements of its Frobenius algebra) and a subgroup  $H'$  which contains no proper normal subgroup of  $G'$ ;  $I = \{i\}$  and  $J = \{j\}$  are two sets of indices and  $w$  is a  $I \times J$  matrix with entries  $w_{ij}$  in  $H'$ . Without loss of generality we can assume that there exists no pair of indices  $j, j' \in J$  ( $i, i' \in I$ ) and no element  $h \in H'$  such that  $w_{ij}h = w_{ij'}(hw_{ij} = w_{ij'})$  identically for all  $i \in I$  ( $j \in J$ ).

We shall denote by  $V$  the set of all  $I \times I$  matrices  $v$  with entries in  $G' \cup \{0\}$  that have the following property: for each  $j \in J$  there exists an index  $j' \in J$  and an element  $g_{jj'} \in G'$  which are such that the product  $vw_{.j}$  (with  $w_{.j}$  = the  $j$ th column vector of  $w$ ) is equal to  $w_{.j'}g_{jj'}$  (i.e., to the vector whose  $i$ th entry is equal to  $w_{ij'}g_{jj'}$ ). Trivially, this condition implies that  $v$  has one and only one non zero entry in each line; it also implies the existence of an isomorphism  $v \rightarrow \bar{v}$  which sends  $V$  onto the monoid  $\bar{V}$  of the  $J \times J$  matrices defined by the symmetric condition and which is such that  $vw = w\bar{v}$ , identically;  $V$  is a monoid and it contains as minimal ideal the set  $V_0$  of all matrices whose  $i$ th column vector is equal to  $w_{.j}g$  (with any  $i \in I, j \in J, g \in G'$ ) and whose  $i'$ th column vector is zero for  $i' \neq i$ .

**IV.1.** The subset  $L \subset V$  of the matrices of  $V$  which have at least one entry in  $H'$  satisfies  $U_r$  and  $U_l$ .

**PROOF.**  $L$  is not empty since it contains at least the neutral element of  $V$ . Let us assume that  $v \in L$  and that  $v_{ii'} \in H'$ . Because of the hypothesis that all the entries of  $w$  belong to  $H'$ , the  $i$ th coordinate of  $vw_{.j}$  for any  $j \in J$ , (that is,  $v_{ii'}w_{i'j}$ ) belongs to  $H'$ . Thus,  $vw_{.j} = w_{.j'}h$  for some  $j' \in J$  and  $h \in H'$ ; it follows that all the non zero entries of  $v$  belong to  $H'$ . This shows that  $L$  is a monoid and, trivially, that it satisfies  $U_r$  and  $U_l$ .

**IV.1'.** If  $F$  is a free monoid and  $\gamma': F \rightarrow V$  an homomorphism, then the subset  $A = \gamma'^{-1}(L \cap \gamma'F)$  is a birecurrent support and the corresponding parameter,  $\delta^*$ , is at most equal to the index of  $H'$  in  $G'$ .

PROOF. The first part of the statement does not need a proof; we verify the second part by showing that for any  $f \in F$  (with the notations of II.2) there exists an injection of  $\Delta Rf$  into the set of the left  $H'$ -cosets. Let  $r, r' \in \Delta Rf$  with, e.g.  $r' = rf'$ ; for any  $i \in I$ , the condition  $(\gamma'r)_{ii'} \neq 0$  defines in a unique manner  $i' \in I$  and  $g = (\gamma'r)_{ii'} \in G'$ . In a similar way, we define  $i'' \in I$  and  $g' \in G'$  by the condition  $0 \neq (\gamma'f')_{i''i''} (= g')$ . Since  $(\gamma'r')_{ii''} = (\gamma'rf')_{ii''} = gg'$  we see that  $g$  and  $g'$  belong to the same  $H'$ -coset if and only if  $g \in H'$ , that is, if and only if  $f' \in A$ , that is, finally, if and only if  $r = r'$  and this ends the proof.

Reciprocally, if  $A$  is a birecurrent support with finite  $\delta^*$  we can take (with the notations of II.5)  $G' = G$  and  $H' = H$  and find,  $I, J$  and  $w$  such that  $\gamma F = M$  is a submonoid of  $V$ . Then  $V_0 \subset \gamma F$  and a sufficient condition that  $\gamma f \in V_0$  is  $\delta f = \delta^*$ . We shall not prove these results here since they are a straightforward application of Clifford's theory [4].

IV.1". If  $\delta^*$  is finite and if for each  $f \in F$  there exists a finite positive  $m$  such that  $\gamma f^m \in V_0$ , then the parameter  $\delta^*$  defined in II.7 is always a divisor of  $\delta^*$ .

PROOF. We consider the group  $G'$  defined in II.8. According to the general theory of monoids [4] the only groups contained in  $\gamma F$  under the hypothesis of IV.1" are in fact contained in  $V_0$ . Consequently, they are isomorphic to subgroups of  $G$  and this concludes the proof.

IV.2. If  $A$  is a birecurrent support such that  $A^*$  is a finite set then either there exists an  $s \in S$  for which  $sF \cap A = \phi$  (and then  $(A, \mu)$  is not persistent for any positive product measure  $\mu$ ) or else, the conditions of IV.1" are satisfied by  $A$ . In this second case,  $\gamma F$  is a group if and only if  $A^*$  reduces to the set of all the words having some fixed finite length. [5].

PROOF. We assume that  $A^*$  is finite and that  $A \cap sF \neq \phi$  for all  $s \in S$ ; then, by the very definition of  $\gamma$  the monoid  $\gamma F$  is finite. By II.2 we see that if  $r, r' \in \Delta Rf$  for some  $f \in F$ , then the equation  $\gamma r = \gamma r'$  implies  $r = r'$ . Thus, the parameter  $\delta^*$  is finite. Let us take any element  $f \in F$ ; the hypothesis that  $\delta f < \delta^*$  implies that for some pair  $(f', f'')$  one has  $f'f'' \in A^*$ . Thus for all  $f \in F$ ,  $\delta f^m = \delta^*$  for large enough  $m$  since, otherwise,  $A^*$  would not be finite. This proves that  $A$  satisfies the conditions of IV.1".

We now make the supplementary assumption that  $\gamma F$  is a group  $G$  with  $\gamma A = H$ , and we consider  $a$ , an element of maximal length of  $A^*$ . If  $|a| = 1$  the result is vacuously true since, then,  $A = F$ . If  $|a| \geq 2$  we write  $a = sxx'$  with  $x, x' \in X$ . Because of  $U_r$ , no left factor of  $a$  belongs to  $A^*$  and because of the maximality of  $|a|$ , we have  $sxx'' \in A$  for all  $x'' \in X$ . Thus, all the generators of  $F$  belong to the same left  $H$ -coset. For this reason, we cannot have  $sx'' \in A^*$  for any  $x'' \in X$  and, because again of the maximal character of  $|a|$  this implies that  $sx''x''' \in A^*$  for any two  $x'', x''' \in X$ . Thus, for any two elements  $x, x' \in X$ , the left coset  $xx'H$  does not depend upon the choice of  $x$  and  $x'$ . If  $|a| = 2$ , this proves the result. If  $|a| \geq 3$  we can write  $s = s'y$  with  $y \in X$  and by the same argument we prove that for any  $x, x', x'' \in X$  the coset  $xx'x''H$  does not depend

upon the choice of these three elements. Since  $|a|$  is finite, by hypothesis, a simple induction gives the result.

The next statements discuss the existence of birecurrent supports with finite  $\delta^*$ . Without loss of generality, we shall assume from now on that  $X$  contains a finite number  $\geq 2$  of elements.

**IV.3.** For any finite  $n \geq 3$  there exist infinitely many different birecurrent supports with this value of  $\delta^*$ .

**PROOF.** In the next section we shall show the existence of at least one birecurrent support with  $\delta^* = 2$  and  $A^*$  infinite. In this section we show that to every birecurrent support  $A$  and element  $u \in A^*$  we can associate one other birecurrent support  $B$  with  $\delta_B^* = \delta^* + 1$  and  $B^*$  infinite and that, for the same  $A^*$  and different choice of  $u \in A^*$ , the two corresponding new supports are different. Thus IV.3. will be entirely proved with the help of IV.4.. Let us now take  $u \in A^*$ , a fixed element, and define:  $J = (uF \cap Fu) - \{u\}$ ;  $J^* = J - J^2$  (i.e., = the subset of those elements of  $J$  that cannot be written as the product of two elements of  $J$ ). With the help of II.1", it is easily verified that there exists a birecurrent support  $B$  which is such that  $B^* = J^* \cup (A^* - \{u\})$  and we prove that for all  $f \in F$  the number (say,  $\delta(B, f)$ ) of its  $B$ -factorizations is at most equal to  $\delta f + 1$ . In order to do this, we slightly extend the notations of II.2, and for any subset  $F'$  of  $F$  we say that the triple  $(f'', f', f''')$  is a  $F'$ -factorization of  $f$  if  $f' \in F'$  and  $f''f''' = f$ ; also, we denote by  $\delta(F', f)$  the number of distinct  $F'$ -factorizations of  $f$  and we observe that by induction on the length of  $f$ , the result of II.3 can be summarized by the identity  $|f| + 1 = \delta(A, f) + \delta(A^*, f)$ .

Here, we have

$$\delta(A^*, f) = \delta(A^* - \{u\}, f) + \delta(\{u\}, f),$$

$$\delta(B^*, f) = \delta(A^* - \{u\}, f) + \delta(J^*, f),$$

We want to show that  $\delta(B^*, f) \leq \delta(A^*, f) + 1$ . If  $\delta(\{u\}, f) = 0$  or 1, we have  $\delta(J^*, f) = 0$  and the result is proved; consequently, we assume now that  $\delta(\{u\}, f) \geq 2$  and we consider two  $\{u\}$ -factorizations  $(f_1, u, f'_1)$  and  $(f_2, u, f'_2)$  with, e.g.  $|f_1| \leq |f_2|$ . The element  $w$  determined by the equation  $f = f_1 w f'_2$  belongs to  $J$ ; it belongs to  $J^*$  if and only if there is no  $\{u\}$ -factorization  $(f_3, u, f'_3)$  for which  $|f_1| < |f_3| < |f_2|$ ; it follows instantly that  $\delta(J^*, f) = \delta(\{u\}, f) - 1$  and the result is proved.

**IV.4.** For each finite  $n \geq 3$  there exist at least two different birecurrent supports with  $A^*$  finite and  $\delta^* = n$ .

**PROOF.** One of these supports has been described in IV.2; in order to produce the other one, we take a birecurrent support  $A$ , a fixed element  $u \in (F - A^*F) \cap (F - FA^*)$  and we construct another birecurrent support  $B$  with  $\delta_B^* = \delta^*$ ; in the last part of the proof we verify that by a proper choice of  $u$  and  $A^*$  we can make  $B^*$  finite.

Let the following sets be defined:

$$C^* = A^* - A^* \cap (uF \cup Fu),$$

$$Z = \{f: uf \in A^* - A^* \cap Fu\},$$

$$Z' = \{f: fu \in A^* - A^* \cap uF\},$$

$$J^* = A^* \cap uF \cap Fu,$$

$$P^* = \{f: fu \in A^* \cap uF\}.$$

Thus,  $A^*$  admits a partition into the sets  $C^*$ ,  $uZ$ ,  $Z'u$  and  $J^*$ ; by construction, there exists a recurrent support  $P$  such that  $P^* = P^2 - P$  (with  $P = \{e\}$  if  $P^*$  is empty) and one can verify that there exists a birecurrent support  $B$  such that  $B^*$  admits a partition into the sets  $C^*$ ,  $\{u\}$  and  $Z'PuZ$ .

In order to verify that  $\delta_b^* = \delta^*$  we take an arbitrary positive product measure  $\mu$  and, for any  $F' \subset F$ , we write  $T(F')$  as an abbreviation for  $\sum (|f| \mu f: f \in F')$ . Thus, by III.3, we have, e.g.,  $\delta^* = T(A, \mu) = T(A^*)$ .

By a simple computation, we obtain when  $\delta^*$  is finite:  $\delta^* = T(A^*) = T(C^*) + T(P^*) + |u|(\mu Z + \mu Z' + \mu P^*)\mu u + (T(Z) + T(Z') + T(P^*))\mu u$ . Also,  $\mu Z = \mu Z' = 1 - \mu P^*$ ;  $\mu P = (1 - \mu P^*)^{-1}$ ;  $T(P) = (1 - \mu P^*)^{-2}T(P^*)$ . Now,  $T(B^*) (= \delta_b^*)$  is equal to the sum  $T(C^*) + |u|\mu u + T(Z'PuZ)$ ; because of the above relations, we have  $T(Z'PuZ) = |u|\mu u\mu Z + (T(Z) + T(Z') + T(P^*))\mu u$  and this concludes the second part of the proof.

Let us now observe that  $B^*$  is finite if and only if  $C^*$  is finite and  $P = \{e\}$ . The first condition is surely satisfied when  $A^*$  is finite and the second one is equivalent to  $P^* = \phi$ , that is, to  $A^* \cap uF \cap Fu = \phi$ .

Thus, if  $A^*$  is the set of all words of length  $n > 2$  and if  $x_1, x_2 \in X$ , the word  $u = x_1^{n-2}x_2$  belongs to  $F - A^*F$  and to  $F - FA^*$  and it satisfies our last condition; this ends the proof of IV.4.

If we take  $n = 2$  and  $u = x_1$  we find that  $P^* = x_1$  and the corresponding  $B^*$  is infinite; this is the example needed for IV.3.

**IV.5.** For each finite  $n$  there exists only a finite number of birecurrent supports  $A$  with  $\delta^* = n$  which satisfy one or the other of the two following supplementary conditions: that  $\gamma F$  is a group or that  $A^*$  is finite.

**PROOF.** This is obvious for the first condition since, because of II.4', it amounts to the fact that for any finite  $n$  there exist only finitely many groups of permutation on  $n$  symbols.

With respect to the second condition we first verify the following elementary remark: let  $K_0 = F - \{e\}$ ,  $K_1, K_2, \dots$  be a decreasing sequence of subsets of  $F$  defined inductively by the relation  $K_{i+1} = \{fFf: f \in K_i\}$ . If  $X$  is finite there exists for every finite  $i$  a finite value  $d(i)$  which is such that every word of length at least  $d(i)$  has at least one factor belonging to  $K_i$ . Indeed, if  $d(i)$  has already been defined, we take  $d(i+1)$  as  $d(i) (1 + |X|^{d(i)})$  where  $|X|$  denotes the number of elements of  $X$ . Then, every word of length  $d(i+1)$  contains at

least two disjoint identical factors of length  $d(i)$  and the result follows by induction.

We now observe that if  $f \neq e$  the hypothesis that  $A^*$  is a finite set (with finite  $\delta^*$ ) implies that  $\delta ff'f \geq \inf(\delta^*, \delta f + 1)$ . Indeed, this is surely true if  $\delta f' > \delta f$  or if  $\delta ff'f = \delta^*$ ; in the remaining case, i.e., in the case that  $\delta f = \delta ff'f < \delta^*$ , we would have according to II.2, for all finite  $m$ ,  $\delta(ff')^m f = \delta f < \delta^*$  and, according to the same remark, there would exist for all finite  $m$  at least one  $a \in A^*$  admitting  $(ff')^m f$  as a factor, which is impossible since  $A^*$  is assumed to be finite.

Thus, by induction, every word  $f$  of length  $\geq d(\delta^*)$  is such that  $\delta f = \delta^*$  and, consequently, it cannot be a factor of a word  $a \in A^*$ . This proves that for given  $\delta^*$  the hypothesis that  $A^*$  is finite imposes that the lengths of the words from  $A^*$  is bounded and it concludes the proof (cf.[6]).

#### REFERENCES

- [1] CHEVALLEY, C., *Fundamental Concepts of Algebra*, Academic Press, New York, 1956.
- [2] FELLER, WILLIAM, *An Introduction to Probability Theory and its Applications*, Vol. 1, 2nd ed., John Wiley and Sons, New York, 1957.
- [3] KESTEN, H., "Symmetric random walks on groups," *Trans. Amer. Math. Soc.*, Vol. 92 (1959), pp. 336-354.
- [4] MILLER, D. D., AND CLIFFORD, A. H., "Regular  $D$ -classes in semigroups," *Trans. Amer. Math. Soc.*, Vol. 82 (1956), pp. 270-280.
- [5] SCHÜTZENBERGER, M. P., "Une théorie algébrique du codage," *Séminaire Dubreil-Pisot*, exp. no 15, Paris, 1955-1956.
- [6] SCHÜTZENBERGER, M. P., "On a family of submonoids." To appear in *Pub. Math. Inst. Hungarian Academy of Science*, December, 1961.

# MAXIMUM LIKELIHOOD CHARACTERIZATION OF DISTRIBUTIONS

BY HENRY TEICHER

*New York University and Purdue University*

**1. Introduction.** It is a commonplace observation that the sample mean and sample variance from a normal population (based on a random sample) are stochastically independent. Considerably less prosaic is the converse proposition, first proved in 1936 by R. C. Geary [2] (under superfluous restrictions), to the effect that the independence of these two statistics entails normality of the underlying population. This, plus the theorem that if two linear combinations (non-zero coefficients) of a pair of independent random variables are themselves independent, the variables are normally distributed, which was proved by Kac in 1939 [4], are harbingers of what are today referred to as characterization theorems. An extensive bibliography of such theorems appears in [5]. Most of these results have the format: if such-and-such statistics are independent (alternatively, if the distribution of such-and-such a statistic is thus-and-so), the underlying population is so-and-so.

The ensuing theorems belong to this genre but adopt a maximum likelihood posture. The first deals with translation (location) parameter and the latter with scale parameter families of distributions.

**2. Preliminaries.** Since the results expounded here concern maximum likelihood estimators, it would seem appropriate to say a few words concerning these. It is somewhat surprising that major treatises on mathematical statistics and estimation do not define maximum likelihood *estimators* per se but merely a maximum likelihood *estimate*. (Pitman's terminological demarcation between these notions will be made explicit shortly.) The definitions of [8], [9] are closest in spirit to that given here.

In order to pave the way for a discussion of these questions, let  $F(x; \theta)$ ,  $-\infty < x < \infty$ ,  $\theta \in \Omega \subset R^1$  denote a one parameter family of probability distributions on the real line  $R^1$  with spectra  $S_\theta$ . Define  $S = \bigcup_{\theta \in \Omega} S_\theta$  and  $S^n = S \times S \times \cdots \times S$ , the  $n$ -fold cartesian product of  $S$  with itself.<sup>1</sup> If, for each  $\theta \in \Omega$ ,  $F(x; \theta)$  is absolutely continuous, designate its probability density function (p.d.f.) by  $f(x; \theta)$ ; if, for each  $\theta$ ,  $F(x; \theta)$  is a step function, the same notation  $f(x; \theta)$  will be used to specify the so-called discrete p.d.f., that is, the mass function of the corresponding distribution (positive at the countable set of points constituting  $S_\theta$  and zero elsewhere).

The customary definition of a maximum likelihood *estimate* of a parameter  $\theta$  of a population (family of distributions generally restricted to the aforementioned types), based on a (random) sample of  $n$  observations  $x_1, x_2, \dots, x_n$ , is a value of  $\theta$ , say  $\hat{\theta}_n$ , which renders  $\prod_{i=1}^n f(x_i; \theta)$  a maximum. A maximum likelihood

Received August 23, 1960; revised January 20, 1961.

<sup>1</sup> When  $\bigcup_{\theta \in \Omega} S_\theta = [a, \infty)$  or  $(-\infty, b]$ , the points  $a, b$  will be deleted in defining  $S$  (so as to avoid a special treatment of the origin in Theorem 2).



estimator (M.L.E.) is presumably a function  $\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \dots, x_n)$  from  $S^n$  into  $\Omega$  which, for every choice of  $x_1, \dots, x_n$ , is a maximum likelihood estimate. (It is by no means apparent that a M.L.E., so defined, is a bona fide random variable and it would seem of interest to give minimal conditions under which it is measurable, [7]. Measurability is, however, tangential to the problem treated here.)

Unfortunately, from the theoretical standpoint, such a definition harbors ambiguities of a petty but disconcerting nature. The fact that these annoyances crop up on sets of measure zero does not seem sufficient reason to ignore their existence.

First, (consider the absolutely continuous case) as a consequence of the fact that for each  $\theta \in \Omega$ ,  $f(x; \theta)$  is defined only to within a set of measure zero, it is possible to change a M.L.E. by altering  $f(x; \theta)$  at one value of  $x$  for each  $\theta$  or even prevent its existence by a perverse choice of  $f(x; \theta)$ . If scale and translation parameter families are involved,  $f(x; \theta)$  is a function of one variable only and the scope for tampering is greatly diminished.

Suppose that a suitable version of  $f(x; \theta)$  has been singled out. Then, a M.L.E.  $\hat{\theta}_n$  will be interpreted as a function from  $S^n$  into  $\Omega$  satisfying

$$(0) \quad \prod_{i=1}^n f(x_i; \hat{\theta}_n) \geq \prod_{i=1}^n f(x_i; \theta)$$

for all  $\theta \in \Omega$  and all  $(x_1, \dots, x_n) \in S^n$ . (Note that if  $R^1 - S$  is non-empty and  $\hat{\theta}_n$  were assigned any value in  $\Omega$  for  $(x_1, \dots, x_n) \in R^n - S^n$ , (0) would hold in the degenerate form  $0 = 0$ .)

Secondly, if all  $(x_1, \dots, x_n)$  in  $S^n$  are pertinent to the definition of a M.L.E., how is one to interpret  $0 \cdot \infty$  if it occurs in (0)? (A p.d.f. may be infinite on a set of measure zero.) The conventional interpretation of this product as zero seems mandatory when the value  $x_0$  for which  $f(x_0; \theta) = 0$  belongs to  $R^1 - S$  (if this set is non-empty) and will be adopted for  $x_0 \in S$  as well.

**3. Characterization theorems.** Theorem 1 which deals with translation parameter families emerges as a generalization and modernization of a result of Gauss [1] when the latter is suitably interpreted and rescued from its context of least squares.<sup>2</sup>

**THEOREM 1:** Let  $\{F(x - \theta), \theta \in R^1\}$  be a translation parameter family of absolutely continuous distributions on the real line and let the version of the p.d.f.  $f(x)$  be lower semi-continuous at  $x = 0$ . If, for all (random) samples of sizes two and three, a maximum likelihood estimate of  $\theta$  is the sample arithmetic mean, then  $F(x)$  is a normal distribution with mean zero.

<sup>2</sup> There is a vast literature consisting of discussions, proofs and reproofs of Gauss' result. In view of the fact that the latter was formulated in a least squares context and further that many notions and distinctions which are today commonplace were then only dimly (if at all) perceived, many of the disquisitions are heuristic and unrigorous by modern standards. Among the multitude of commentaries, three ([10], [11], [12] p. 169) are singled out for reference.

All prior proofs which have come to the writer's attention assume implicitly or explicitly that the density function  $f(x)$  is differentiable.

PROOF: If  $\bar{x} = (n+1)^{-1} \sum_{i=1}^{n+1} x_i$ , it follows from (0) and the hypothesis that for  $n = 1, 2$  and all real  $x_1, x_2, \dots, x_{n+1}$  and  $\theta$ ,

$$\prod_{i=1}^{n+1} f(x_i - \bar{x}) \geq \prod_{i=1}^{n+1} f(x_i - \theta);$$

hence,

$$(0)' \quad \prod_{i=1}^{n+1} f(y_i) \geq \prod_{i=1}^{n+1} f(y_i - \theta)$$

for  $n = 1, 2$  and all real  $\theta$  and  $y_1, \dots, y_{n+1}$  satisfying  $\sum_{i=1}^{n+1} y_i = 0$ .

Set  $n = 1, y_1 = y = -y_2$  in (0)' to obtain

$$(1) \quad f(y)f(-y) \geq f(y - \theta)f(-y - \theta), \quad \text{all real } y, \theta.$$

Note that  $f(0) = 0$  implies  $f(y)$  vanishes identically. Suppose that  $f(a) = \infty$  for some real number  $a$ . Then, according to (1), for each  $y \in R^1$  either  $f(y)f(-y) = \infty$  or  $f(2y + a) = 0$ . For a p.d.f., the former cannot hold on a set of positive (Lebesgue) measure, while the latter cannot hold almost everywhere. Thus, every p.d.f. satisfying (0)' is positive at the origin and everywhere finite (so the product  $0 \cdot \infty$  will not arise in (0)').

Let  $h(x) = \log_e f(x)$  where  $h(x)$  may possibly assume the extended real value  $-\infty$ . Then it follows from (0)' that for  $n = 1, 2$  and all real  $y_1, y_2, \dots, y_n, \theta$ ,

$$(2) \quad \sum_{i=1}^n h(y_i) + h\left(-\sum_{i=1}^n y_i\right) \geq \sum_{i=1}^n h(y_i - \theta) + h\left(-\sum_{i=1}^n y_i - \theta\right).$$

As will be seen, under the meager assumptions contained in the probabilistic framework, the functional inequality (2) determines  $h(x)$  and therefore  $f(x)$ .

In particular, (2) implies

$$nh(y) + h(-ny) \geq nh(y - \theta) + h(-ny - \theta),$$

which, for  $n = 1$ , becomes

$$(3) \quad h(y) + h(-y) \geq h(y - \theta) + h(-y - \theta).$$

Note that if in (2),  $\theta$  is replaced by  $-\theta$  and  $y_i$  by  $-y_i$ , the resulting inequality when added to (2) reveals that, if  $g(y)$  is a solution of (2), so is  $h(y) = g(y) + g(-y)$  and we therefore confine attention at first to symmetric solutions of (2), which then takes the form

$$(2)' \quad \sum_{i=1}^n h(y_i) + h\left(\sum_{i=1}^n y_i\right) \geq \sum_{i=1}^n h(y_i - \theta) + h\left(\sum_{i=1}^n y_i + \theta\right).$$

Similarly, (3) becomes

$$(3)' \quad 2h(y) \geq h(y - \theta) + h(y + \theta), \quad \text{all } y, \theta.$$

Suppose that  $h(y) = -\infty$  for some  $y > 0$  and let  $c$  be the infimum of the

set  $A$  of all such positive values  $y$ . Taking  $n = 2$ ,  $y_1 = y_2 = \frac{1}{2}c_m$ ,  $\theta = -\frac{1}{2}c_m$  in (2)' yields  $2h(\frac{1}{2}c_m) + h(c_m) \geq 3h(\frac{3}{2}c_m)$ . Choose  $c_m \searrow c$ ,  $c_m \in A$ ,  $m = 1, 2, \dots$ , implying  $h(\frac{3}{2}c_m) = -\infty$ ,  $m \geq 1$ . If  $c > 0$ , a contradiction ensues, while if  $c = 0$ ,  $f(x)$  is not lower semi-continuous at zero.

Thus,  $h(y)$  is everywhere finite and according to (3)' concave. Since any p.d.f.  $f(x)$  is necessarily measurable,  $h(x)$  is likewise, whence [3, 6],  $h$  is a continuous concave function.

Let  $D$  denote the complement of the (at most countable) set of points at which  $h(x)$  is not differentiable and denote by  $q(x)$  the derivative of  $h(x)$ . Then  $q(x)$  is monotone and defined for all  $x \in D$ .

The fact, as expressed by (2)', that  $\theta = 0$  maximizes  $\sum_{i=1}^n h(y_i - \theta) + h(\sum_{i=1}^n y_i + \theta)$  now requires that

$$(4)' \quad -\sum_{i=1}^n q(y_i) + q\left(\sum_{i=1}^n y_i\right) = 0$$

for all  $y_i \in D$  such that  $\sum_{i=1}^n y_i \in D$ ,  $n = 1, 2$ . For  $n = 2$ , (4)' becomes

$$(5)' \quad q(y_1) + q(y_2) = q(y_1 + y_2) \quad \text{for } y_1, y_2, y_1 + y_2 \in D.$$

Let  $C = \{f\}$  be the class of non-negative measurable functions on  $R^1$  which are everywhere finite, lower semi-continuous at zero, and do not vanish almost everywhere. Let  $C'$  be the subclass of functions in  $C$  which do not vanish anywhere. Since the only monotone solution of Cauchy's functional equation (5)' is  $q(y) = c_1 y$ , it follows that the only symmetric functions of  $C$  satisfying (0)' (which are necessarily in  $C'$ ) are given by  $h(y) = \log_e f(y) = -cy^2 + d$  for  $y \in D$  and therefore by continuity for all real  $y$ .

Suppose next that  $f(y)$  is any element of  $C$  satisfying (0)'. According to (1),  $f(y) \cdot f(-y)$  does not vanish almost everywhere (take  $\theta = -y$ ); it is readily seen that  $f(y) \cdot f(-y)$  is a symmetric function in  $C$  and, as previously noted, a solution of (0)'. Thus  $f(y) \cdot f(-y) \in C'$  implying  $f(y) \in C'$ .

In fact, necessarily for some real constants  $c$  and  $d$ ,

$$g(y) = \log_e f(y) = -\frac{1}{2}(cy^2 - d) + b(y)$$

where  $b(y)$  is an odd function. For, by the preceding,  $g(y) + g(-y) = -cy^2 + d$  for some  $c, d$  and this implies that  $b(y) = g(y) + \frac{1}{2}(cy^2 - d)$  satisfies  $b(y) = -b(-y)$ .

Substituting for  $g(y)$  in (3) yields

$$(4) \quad c\theta^2 \geq b(y - \theta) - b(y + \theta), \quad \text{all } y, \theta.$$

Replacing  $y$  by  $-y$  and  $\theta$  by  $-\theta$  in (4) and combining the result with (4), produces

$$(5) \quad |b(y - \theta) - b(y + \theta)| \leq c\theta^2, \quad \text{all } y, \theta,$$

which, in turn, necessitates  $c \geq 0$  and implies that  $b(y)$  is differentiable and constant, hence identically zero.

Consequently, the only solutions of  $(0)'$  in  $C$  are given implicitly by  $h(x) = -\frac{1}{2}cx^2 + d$ ,  $c \geq 0$  and thus the only p.d.f.'s satisfying the conditions of the theorem are  $f(x) = (c/2\pi)^{1/2}e^{-\frac{1}{2}cx^2}$ , that is, normal density functions with mean zero.

REMARK 1: The integers two and three of the theorem may clearly be replaced by other pairs, e.g.,  $2k, 3k$ ,  $k > 1$ . It seems most desirable, however, to state the result with minimal  $n$ .

REMARK 2: If  $\int xf(x) dx$  exists and is zero, the translation parameter  $\theta$  is the mean of the distribution  $F(x - \theta)$ . In such cases (excluding the normal), the theorem implies that the sample mean is not (for samples of sizes both two and three, a fortiori, for all  $n$ ) a maximum likelihood estimator of the population mean. This is readily seen, for example, if

$$f(x; \theta) = C_\alpha \cdot \exp \{-|x - \theta|^\alpha\}, \quad \alpha \neq 2.$$

REMARK 3: It seems of interest to note in the case where  $F(x - \theta)$  is a rectangular distribution with mean  $\theta$ , i.e.,  $f(x) = 1$  for  $|x - \theta| \leq \frac{1}{2}$  and zero otherwise, that, whereas  $\bar{x}$  is a M.L.E. of  $\theta$  for  $n = 2$ , its numerical value is not a maximum likelihood estimate of  $\theta$  for all random samples of size three.

**4. Scale parameter families.** Consider a scale parameter family<sup>2</sup> of absolutely continuous distributions  $\mathfrak{F} = \{F(x/\sigma), \sigma > 0\}$ . The joint density function of  $n$  independent random variables, each distributed as  $F(x/\sigma)$ , is  $\sigma^{-n} \prod_{i=1}^n f(x_i/\sigma)$  where  $F(x) = \int_{-\infty}^x f(u) du$ . To say that  $\hat{\sigma} = \hat{\sigma}(x_1, \dots, x_n)$  is a maximum likelihood estimator of  $\sigma$  is to say for all  $\sigma > 0$  and  $x_1, \dots, x_n$  in  $S^n$  that

$$\hat{\sigma}^{-n} \prod_{i=1}^n f(x_i/\hat{\sigma}) \geq \sigma^{-n} \prod_{i=1}^n f(x_i/\sigma).$$

Let  $y_i = x_i/\hat{\sigma}$ ,  $\lambda = \hat{\sigma}/\sigma$ . Then, if  $\hat{\sigma}$  is a homogeneous function of degree one in  $x_1, \dots, x_n$ , the preceding implies

$$(6) \quad \prod_{i=1}^n f(y_i) \geq \lambda^n \prod_{i=1}^n f(\lambda y_i)$$

for all  $\lambda > 0$  and  $y_1, \dots, y_n$  satisfying

$$(7) \quad \hat{\sigma}(y_1, y_2, \dots, y_n) = 1.$$

If  $h(y) = \log_e f(y)$  is finite valued, (6) may be transcribed as

$$(8) \quad \frac{1}{n} \sum_{i=1}^n [h(y_i) - h(\lambda y_i)] \geq \log_e \lambda.$$

Inspection shows that  $h(y) = -\log_e y + \text{const.}$  satisfies (8), with equality holding for all choices of  $y_1, \dots, y_n$  and a fortiori for  $y_i$ 's satisfying (7). How-

<sup>2</sup> The standard device of reducing a scale parameter family to a translation parameter family (so as to utilize Theorem 1) appears fruitless here.

ever,  $f(y) = cy^{-1}$  is not integrable on  $(0, \infty)$  and truncation to a finite interval will be precluded by the conditions of Theorems 2 and 3.

In the following theorems the indispensable absolute continuity assumption is augmented by a possibly dispensable continuity assumption of the density function. The seemingly ad hoc condition (ii) on the other hand, appears to be crucial.

**THEOREM 2:** Let  $\{F(x/\sigma), \sigma > 0\}$  constitute a scale parameter family of absolutely continuous distributions with the version of the p.d.f.  $f(x)$  satisfying

$$(i) \quad f(x) \text{ is continuous on } (0, \infty)$$

$$(ii)^4 \quad \lim_{y \rightarrow 0} \frac{f(\lambda y)}{f(y)} = 1, \quad \text{all } \lambda > 0.$$

If, for all sample sizes, a maximum likelihood estimator of  $\sigma$  is the sample arithmetic mean, then  $F$  is the exponential distribution, i.e.,

$$f(x) = e^{-x}, \quad x > 0, \quad f(x) = 0, \quad x \leq 0.$$

**PROOF:** Since  $\Omega = \{\sigma: \sigma > 0\}$  and  $\bar{x}$  is a posited M.L.E., necessarily  $S \subset S' = \{x: x > 0\}$ , whence  $f(x) \equiv 0$  in  $R^1 - S'$ . It suffices, therefore, to consider  $f(x)$ ,  $x \in S'$ , noting that  $f(x) \neq 0$  in  $S'$ . Infinite values of  $f$  are precluded by continuity and it will now be shown that  $f(x) > 0$  in  $S'$ , i.e.,  $S = S'$ .

From prior remarks, (6) obtains with (7) becoming

$$(7.1) \quad \sum_{i=1}^n y_i = n.$$

In (6), choose  $y_i = k/m$ ,  $1 \leq i \leq m$  and  $y_i = [(n-k)/(n-m)]$ ,  $m+1 \leq i \leq n$ , where  $k, m, n$  are positive integers satisfying  $k < m < n$ ; this yields

$$(9) \quad f^m\left(\frac{k}{m}\right) f^{n-m}\left(\frac{n-k}{n-m}\right) \geq \lambda^m f^m\left(\frac{\lambda k}{m}\right) f^{n-m}\left(\frac{\lambda(n-k)}{n-m}\right).$$

Let  $k/m \rightarrow \alpha$ ,  $m/n \rightarrow c$ . Then for all positive  $\lambda$  and all  $c, \alpha$  in  $(0, 1)$

$$(10) \quad f^c(\alpha) f^{1-c}\left(\frac{1-\alpha c}{1-c}\right) \geq \lambda^c f^c(\lambda \alpha) f^{1-c}\left(\frac{\lambda(1-\alpha c)}{1-c}\right).$$

Now, if there exists a sequence  $\alpha_n \rightarrow 0$  with  $f(\alpha_n) > 0$ ,  $n = 1, 2, \dots$ , it follows from (10) and (ii) that

$$(11) \quad f(y) \geq \lambda^2 f(\lambda y) \quad \text{for } y \geq 1, \quad \lambda > 0.$$

Thus, if  $f$  vanished for some  $y \geq 1$ , it would vanish identically. Further, from (6), if  $f(y)$  were zero for some  $y$  in  $(0, 1)$ ,  $f(y)$  would have zeros in  $(1, \infty)$ .

Alternatively, suppose that for some  $\delta > 0$ ,  $f(y) \equiv 0$  in  $(0, \delta)$ . Since  $f(y) \neq 0$

<sup>4</sup> This condition is automatically fulfilled if  $0 < \lim_{x \rightarrow 0} f(x) < \infty$ . Also, it is reiterated that only random samples are under consideration.

in  $S'$ , (6) insures ( $y_i \equiv 1$ ) that  $f(1) > 0$ , whence  $\delta < 1$  and there is no loss of generality in supposing that  $(0, \delta)$  is the maximal interval in which  $f$  vanishes identically. Then, from (10) follows

$$(12) \quad 0 = \lambda f^c(\lambda\delta) f^{1-c} \left( \frac{\lambda(1-\delta c)}{1-c} \right)$$

for all  $\lambda > 0$  and  $c$  in  $(0, 1)$ . By continuity,  $f(y) > 0$  in  $(1-\epsilon, 1+\epsilon)$  for all sufficiently small  $\epsilon > 0$ . Hence, taking  $\lambda = (1-\epsilon)/\delta$  in (12),

$$f\{[(1-\epsilon)(1-\delta c)]/\delta(1-c)\} = 0$$

for  $0 < c < 1$  implying  $f[(1-\epsilon)/\delta] = 0$ , all sufficiently small  $\epsilon > 0$ .

Now, let  $k+1$  be an integer greater than  $(1-\epsilon)/\delta$ . From (6),

$$0 \geq f[\lambda(1-\epsilon)/\delta] \cdot f^k\{\lambda[k+1-(1-\epsilon)\delta^{-1}]/k\}.$$

Hence,  $f\{[\delta/k(1-\epsilon)][k+1-(1-\epsilon)\delta^{-1}]\} = 0$  for all sufficiently large  $k$  and all sufficiently small  $\epsilon$ , implying  $f[\delta/(1-\epsilon)] = 0$  for all sufficiently small  $\epsilon > 0$ , which contradicts the maximality of  $\delta$ .

Thus, any p.d.f. satisfying (6) is non-zero in  $S'$ . Consequently, (11) holds unconditionally and may be rewritten as

$$(13) \quad y^{-1}[h(y) - h(\lambda y)] \geq \log_e \lambda, \quad y \geq 1, \lambda > 0$$

where, as before,  $h(y) = \log_e f(y)$ .

Replace  $\lambda$  by  $\lambda^{-1}$  in (13) and combine the result with (13) to obtain

$$(14) \quad 0 \geq h(\lambda y) - 2h(y) + h(y/\lambda), \quad y \geq 1, \lambda > 0.$$

This asserts that  $H(y) = h(e^y)$  is concave for  $y \geq 0$  and hence that  $h(y)$  is differentiable in  $(1, \infty)$  except perhaps for a countable subset  $D$  thereof. From (13), for  $\lambda < 1$  and  $y \geq 1$ ,

$$\frac{h(\lambda y) - h(y)}{y(\lambda - 1)} \geq \frac{\log_e \lambda}{1 - \lambda} \geq \frac{h(y/\lambda) - h(y)}{\lambda y[(1/\lambda) - 1]}$$

whence  $(\lambda \nearrow 1)$ ,  $h'(y) = -1$  on  $(1, \infty) - D$ . Then by continuity,

$$(15) \quad h(y) = -y + c, \quad y \geq 1.$$

Next, choosing  $y_i < 1$ ,  $i = 1, \dots, r < n$  and  $y_j > 1$ ,  $j > r$ , in (6) and employing (15) and (7.1), we find for all  $\lambda > 0$ ,  $r < n$  and  $y_i$  in  $(0, 1)$  that

$$(16) \quad \sum_{i=1}^r [h(y_i) - h(\lambda y_i) + (1-\lambda)y_i] \geq n[\log_e \lambda + 1 - \lambda].$$

For  $0 < y < 1$ , (16) asserts ( $r = 1$ ) that

$$(17) \quad 1/n[h(y) - h(\lambda y)] \geq \log_e \lambda + (1-\lambda)(1-(y/n)).$$

But for  $0 < x < 1$ ,  $\log_e \lambda + x(1-\lambda) > 0$  for all  $\lambda$  sufficiently close to and larger than unity. Thus, from (17),  $h$  is monotone decreasing in  $(0, 1)$ .



Set  $h = h_1 + h_2$  where  $h_2$  is absolutely continuous and  $h_1$  is singular, i.e.,  $h'_1(y) = 0$  a.e. in  $(0, 1)$  and  $h_1(y) \equiv 0$  for  $y > 1$ .

Again taking  $r = 1$  in (16), there follows for  $0 < y < 1$  and  $\lambda < 1$

$$(18) \quad \frac{h(\lambda y) - h(y) + (\lambda - 1)y}{\lambda - 1} \geq n \left[ \frac{\log_e \lambda}{1 - \lambda} + 1 \right].$$

This implies  $y[h'_2(y) + 1] \geq 0$  almost everywhere in  $(0, 1)$ . Similarly, replacing  $\lambda$  by  $\lambda^{-1}$  in (16), the inequality is reversed. Hence,  $h'_2(y) = -1$ , almost everywhere in  $(0, 1)$  implying  $h_2(y) = -y + c_1$ .

Utilizing this result in (18), there follows

$$(19) \quad \limsup_{\lambda \rightarrow 1} \frac{h_1(\lambda y) - h_1(y)}{y(\lambda - 1)} \geq \limsup_{\lambda \rightarrow 1} \frac{h_1(\lambda y) - h_1(y)}{y(\lambda - 1)} \geq 0$$

for all  $y$  in  $(0, 1)$ ; hence  $h_1(y) \equiv 0$ . Also, by continuity  $c_1 = c$ .

Thus, for  $y \in S'$ ,  $f(y) = ae^{-y}$  and since  $f$  vanishes outside  $S'$ ,  $a = 1$ .

THEOREM 3: Let  $\{F(x/\sigma), \sigma > 0\}$  be a scale parameter family of absolutely continuous distributions with the version of the p.d.f.  $f(x)$  satisfying

$$(i) \quad f(x) \text{ continuous on } (-\infty, \infty)$$

$$(ii)^4 \quad \lim_{y \rightarrow 0} [f(\lambda y)/f(y)] = 1, \quad \text{all } \lambda > 0$$

If, for all sample sizes, a maximum likelihood estimator of  $\sigma$  is  $(n^{-1} \sum_{i=1}^n x_i^2)^{1/2}$ , then  $F(x)$  is the normal distribution with mean zero and variance one.

PROOF: Here (7) specializes to

$$(7.2) \quad \sum_{i=1}^n y_i^2 = n$$

In (6), set  $y_i = \pm(k/m)^{1/2}$ ,  $1 \leq i \leq m$ ;  $y_i = \pm[(n-k)/(n-m)]^{1/2}$ ,  $m+1 \leq i \leq n$ . Analogous to (10), there follows

$$(10)' \quad f^c(\alpha)f^{1-c} \left( \pm \sqrt{\frac{(1-\alpha^2c)}{(1-c)}} \right) \geq \lambda f^c(\lambda\alpha)f^{1-c} \left( \pm \lambda \sqrt{\frac{(1-\alpha^2c)}{(1-c)}} \right)$$

valid for  $\lambda > 0$ ,  $|\alpha| \leq 1$ ,  $0 < c < 1$ . An argument akin to that employed in Theorem 2 shows that  $f$  is non-vanishing and it follows from (10)' that

$$y^{-2}[h(y) - h(\lambda y)] \geq \log_e \lambda, \quad |y| \geq 1, \lambda > 0.$$

Again,  $h$  is differentiable, this time in the region  $A: |y| \geq 1$  except perhaps on a countable subset  $D'$  thereof. Proceeding as in the proof of Theorem 2, we find that  $h'(y) = -y$  for  $y$  in  $A - D'$  and hence that  $h(y) = -\frac{1}{2}y^2 + c$  for  $|y| \geq 1$ . The analogue of (16) is

$$(16)' \quad \sum_{i=1}^{i=r} [h(y_i) - h(\lambda y_i) + (1 - \lambda^2)\frac{1}{2}y_i^2] \geq n [\log_e \lambda + \frac{1}{2}(1 - \lambda^2)]$$

where  $|y_i| < 1$  and  $r < n$ . This implies that  $h$  is decreasing in  $(0, 1)$  and increas-

ing  $a(-1, 0)$ . An argument paralleling that of Theorem 2 yields  $h(y) = -\frac{1}{2}y^2 + c$ , for  $|y| < 1$  implying  $f(y) = a \exp\{-y^2/2\}$ . Finally,  $a = (2\pi)^{-1}$ .

## REFERENCES

- [1] CARL FRIEDRICH GAUSS, "Theoria motus corporum coelestium," *Werke*, liber II, Sectio III, pp. 240-44. See also "Gauss' work on the theory of least squares" by Hale F. Trotter, pp. 127-33, Statistical Research Group, Princeton, N. J.
- [2] R. C. GEARY, "Distribution of Student's ratio for non-normal samples," *J. Roy. Stat. Soc. Suppl.*, Ser. B, Vol. 3 (1936), pp. 178-84.
- [3] H. BLUMBERG, "On convex functions," *Trans. Amer. Math. Soc.*, Vol. 20 (1919), pp. 40-44.
- [4] M. KAC, "On a characterization of the normal distribution," *Amer. J. Math.*, Vol. 61 (1939), pp. 726-28.
- [5] E. LUKACS, "Characterization of populations by properties of suitable statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1956, pp. 195-214.
- [6] W. SIERPINSKI, "Sur les fonctions convexes mesurables," *Fundamenta Math.*, Vol. 1 (1920), pp. 125-9.
- [7] R. R. BAHADUR, "On the asymptotic efficiency of tests," *Sankhyā*, Vol. 22 (1960), pp. 229-52.
- [8] J. HANNAN, "Consistency of maximum likelihood estimation of discrete distributions," *Contributions to Probability and Statistics*, Stanford University Press, Stanford, Calif., 1960, pp. 249-57.
- [9] LUCIEN LECAM, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *University of California Publications in Statistics*, Vol. 1, pp. 277-330.
- [10] J. GLAISHER, "On the law of facility of errors of observations and on the method of least squares," *Mem. Astron. Soc.*, London, Vol. 39 (1872), pp. 75-124.
- [11] MANSFIELD MERRIMAN, "A list of writings relating to the method of least squares with historical and critical notes," *Trans. Connecticut Academy of Arts and Sciences*, Vol. 4 (1877-82), pp. 151-231.
- [12] H. POINCARÉ, *Calcul des Probabilités*, Gauthier-Villars, Paris, 1912.

## ON THE DISTRIBUTION OF FIRST SIGNIFICANT DIGITS<sup>1</sup>

BY ROGER S. PINKHAM

*Rutgers—The State University*

**Introduction.** It has been noticed by astute observers that well used tables of logarithms are invariably dirtier at the front than at the back. Upon reflection one is led to inquire whether there are more physical constants with low order first significant digits than high. Actual counts by Benford [2] show that not only is this the case but that it seems to be an empirical truth that whenever one has a large body of physical data, Farmer's Almanac, Census Reports, Chemical Rubber Handbook, etc., the proportion of these data with first significant digit  $n$  or less is approximately  $\log_{10}(n + 1)$ . Any reader formerly unaware of this "peculiarity" will find an actual sampling experiment wondrously tantalizing. Thus, for example, approximately 0.7 of the physical constants in the Chemical Rubber Handbook begin with 4 or less ( $\log_{10}(4 + 1) = 0.699$ ). This is to be contrasted with the widespread intuitive evaluation  $\frac{4}{10}$ ths.

At least two books call attention to this peculiarity, Furlan [6] and Wallis [18], but to my knowledge there are only five published papers on the subject, Benford [2], Furry et al [7], [9], Gini [8], and Herzel [11]. The first consists of excellent empirical verifications and a discussion of the implied distribution of 2nd, 3rd, ... significant digits. The second and third put forth the thesis that the distribution of significant digits should not depend markedly on the underlying distribution, and the authors present numerical evaluations for a range of underlying distributions in support of their contention. The fourth maintains that explanation is to be sought in empiric considerations. The fifth considers three different urn models; each yields a distribution of initial digits which the author compares with  $\log_{10}(n + 1)$ .

This paper is a theoretical discussion of why and to what extent this so called "abnormal law" must hold. The flavor of the results is, I think, conveyed in the following remarks.

(i) The only distribution for first significant digits which is invariant under scale change of the underlying distribution is  $\log_{10}(n + 1)$ . Contrary to suspicion this is a non-trivial mathematical result, for the variable  $n$  is discrete.

(ii) Suppose one has a horizontal circular disc of unit circumference which is pivoted at the center. Let the disc be given a random angular displacement  $\theta$  where  $-\infty < \theta < \infty$ . If the final position of the disc mod one is called  $\varphi$ , i.e.,

$$\varphi \equiv \theta \bmod(1), \quad 0 \leq \varphi < 1,$$

---

Received June 23, 1960; revised June 12, 1961.

<sup>1</sup> This work was supported by the Office of Naval Research under Contract Nonr 404 (16). Reproduction in whole or in part is permitted for any purpose of the United States Government.

then  $\varphi$  is a random variable whose probability structure is determined entirely by that of  $\theta$ . In fact if

$$\Pr(x \leq \theta < x + dx) = g(x) dx,$$

and

$$\Pr(y \leq \varphi < y + dy) = f(y) dy,$$

then

$$f(y) = \sum_{m=-\infty}^{\infty} g(y + m).$$

Now it is intuitively obvious that for a wide range of possible distributions of  $\theta$  the distribution of  $\varphi$  should be approximately uniform i.e.,

$$f(y) \approx 1, \quad 0 \leq y \leq 1.$$

This and related properties of distributions wrapped around a circle have been known for some time, Dvoretzky [4], Lévy [14], Robbins [15], and put to various uses, Aitchison [1], Brown [3], Horton and Smith [12], Tochar [17].

The logarithmic law of left-most significant digits is a consequence of the above property of random variables mod one. One can see this as follows. Let  $F(x)$  be the cumulative distribution function for the population of physical constants (taken non-negative for convenience). Define  $D(x)$  by

$$D(x) = \sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)], \quad x > 0.$$

$D(n)$  for  $n = 2, 3, \dots, 10$  gives the proportion of the population with first significant digit  $n - 1$  or less. The logarithmic "law" states that  $D(n)$  should be approximately  $\log_{10}(n)$ . Thus one suspects that

$$\log_{10}(x) \approx \sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)].$$

A change of variables will make clear the connexion with the spinning disc. Let

$$y = \log_{10}(x) \text{ and } G(y) = F(10^y).$$

One then has

$$y \approx \sum_{m=-\infty}^{\infty} [G(y + m) - G(m)],$$

or, taking derivatives,

$$1 \approx \sum_{m=-\infty}^{\infty} g(y + m).$$

This latter approximate equality is the one mentioned before in connexion with random variables mod one.

Section 1 gives the mathematical support for contention (i) while Section 2 provides a mathematical basis for the approximation alluded to in (ii).

After the mathematical work of Section 2 had been completed I discovered the basic mathematical idea without the detail in a discussion by I. J. Good of a paper by Tocher [17].

**1. An invariance principle.** The population of known physical constants changes daily, but the collection of such constants can be regarded as a large sample from an unknown underlying distribution of all physical constants. It is this underlying distribution in which interest will center.

Such mental constructs are familiar in the natural sciences. Thus most physical objects are regarded as having a density even though they are "known" to have a granular structure at the atomic level. Such entities are of course outside the compass of mathematics *per se*.

Consider the population of all physical constants and the derived distribution of first significant digits. Suppose all the physical constants were multiplied by some fixed number. What would happen to the distribution of first significant digits? One feels, I think, that it would be the same as before. This invariance property is enough, as is shown below, to characterize the distribution completely.  $\log_{10}(n + 1)$  emerges as the necessary cumulative. The basic mathematical fact is that a certain derived functional equation has one and only one solution.

Suppose  $F(x)$  is the cumulative distribution function for the population of all physical constants (assumed non-negative) in accordance with their size. Then

$$(1) \quad D(x) = \sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)], \quad x > 0,$$

is a well defined function for positive  $x$ ;  $D(n)$  for  $n = 2, \dots, 9, 10$  gives the proportion with first significant digit  $n - 1$  or less, since all numbers between  $10^m$  and  $n \times 10^m$  begin with  $n - 1$  or less.

If all the physical constants are multiplied by a positive constant  $c$ , then the resulting cumulative is  $F(x/c)$ . The postulated invariance yields

$$D(n) = \sum_{m=-\infty}^{\infty} \left[ F\left(\frac{n}{c} 10^m\right) - F\left(\frac{10^m}{c}\right) \right]$$

or

$$(2) \quad D(n) = D(n/c) - D(1/c), \quad c > 0; n = 2, \dots, 10.$$

If the relation (2) held for arbitrary positive real  $n$  rather than  $n = 2, \dots, 10$  one could, assuming continuity, immediately deduce  $D(n) = \log_{10} n$ . We now show this conclusion to be justified under even weaker conditions than are implicit in Equation (2).

THEOREM 1. If

1.  $D(2) + D(x) = D(2x), \quad x > 0;$
2.  $D(10) + D(x) = D(10x), \quad x > 0;$
3.  $D(x)$  is continuous;
4.  $D(10) = 1;$

then  $D(x) = \log_{10}(x), x > 0$ .

PROOF. Let  $H(x) = D(10^x)$ . Then conditions 1 and 2 become

$$(3) \quad H(\log n) + H(y) = H(\log n + y), \quad -\infty < y < \infty, n = 2, 10.$$

Thus  $H(N \log n) = NH(\log n)$  if  $N$  integral, and one has  $H(N) = N$  since  $H(1) = 1$ . From the theory of continued fractions one knows, Hardy and Wright [10],

$$\log 2 = (p_m/q_m) + o(1/q_m) \quad (m \rightarrow \infty)$$

with  $p_m, q_m$  integers. I.e.,

$$q_m \log 2 = p_m + o(1) \quad (m \rightarrow \infty).$$

Hence by hypothesis 2

$$q_m H(\log 2) = p_m + o(1) \quad (m \rightarrow \infty).$$

Therefore  $H(\log 2) = \log 2$ . Suppose  $a$  irrational, and let  $[x]$  denote the largest integer not exceeding  $x$ . Then it is well known, Kac([13], p. 41), that the sequence

$$a_n = na - [na], \quad n = 1, 2, \dots,$$

is uniformly distributed on  $[0, 1]$ . Thus there exists a subsequence  $a_{n'}$  converging to any fixed  $h$  ( $0 \leq h < 1$ ). Take  $a$  to be  $\log 2$ .

$$H(a_{n'}) = n'H(\log 2) - [n'\log 2] = a_{n'}.$$

Letting  $n'$  tend to infinity yields, by the assumed continuity,  $H(h) = h$ . Since  $y = [y] + y - [y]$ ,  $H(y) = y$ , ( $-\infty < y < \infty$ ), and  $D(x) = \log_{10} x$ .

It is reasonable to consider  $F(x)$  continuous from which it follows that the  $D(x)$  of (1) is continuous and thence by the Theorem 1 that  $D(x) = \log_{10}(x)$ .

**2. An approximation.** Drop from consideration any invariance postulate. Consider to what degree  $\log_{10} x$  provides an approximation to

$$\sum_{m=-\infty}^{\infty} [F(x10^m) - F(10^m)], \quad 1 \leq x \leq 10.$$

( $F(\cdot)$  has the same significance as before.) Let  $G(y) = F(10^y)$ . Then one may as well consider how  $x$  approximates

$$(4) \quad J(x) = \sum_{m=-\infty}^{\infty} [G(x+m) - G(m)], \quad 0 \leq x \leq 1.$$



It is reasonable to take some canonical representation of  $G(x)$  and hope that the sum  $J(x)$  can be evaluated explicitly. A statistician immediately thinks of Fourier transforms since characteristic functions always exist for distributions.

A trouble immediately appears. In trying to evaluate the sum

$$J(x) = \sum_{m=-\infty}^{\infty} [G(x+m) - G(m)],$$

one immediately stubs against sums of the form

$$\sum_{m=0}^{\infty} e^{imx},$$

which of course do not converge. To overcome this difficulty one may introduce a "convergence factor" and subsequently sneak it out again at the end.

Thus, define  $J(x|t)$  by

$$(5) \quad J(x|t) = \sum_{m=-\infty}^{\infty} [G(x+m) - G(m)]t^{|m|}, \quad 0 < t < 1,$$

and  $W(u)$  by  $W(u) = \int_{-\infty}^{\infty} e^{iut} dG(t)$ . Then  $W(u)$  exists for all  $-\infty < u < \infty$ , and

$$G(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-ixu}}{iu} W(u) du.$$

Suppose

$$W(u) = O(u^{-h}), \quad h > 0, \quad (|u| \rightarrow \infty)$$

Then, by merely summing geometric series after switching the order of summation and integration, one has

$$J(x|t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-ixu}}{iu} W(u) P(u, t) du;$$

$P(u, t)$  is the Poisson kernel given by

$$(6) \quad P(u, t) = \frac{1 - t^2}{1 + t^2 - 2t \cos u}.$$

The interchange of limits is justified by the assumed order condition.

The Poisson kernel when properly normalized is a frequency function uncommon in statistical circles. Thus

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} P(u, t) du = 1 \quad \text{and} \quad P(-u, t) = P(u, t).$$

Furthermore, the variance of the distribution goes to zero as  $t$  tends to one. Hence

$$\lim_{t \rightarrow 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} P(u, t) Q(u) du = Q(0),$$

if  $Q$  is continuous at the origin.

Now to return to  $J(x|t)$ . Splitting the integral up into integrals over contiguous intervals of length  $2\pi$ , and utilizing the periodicity of  $P(u, t)$  yields

$$J(x|t) = \sum_{k=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1 - e^{-ix(u+2\pi k)}}{iu + i2\pi k} W(u + 2\pi k) P(u, t) du.$$

A second appeal to the assumed order conditions on  $W(u)$  allows one to take limits as  $t \rightarrow 1$  term by term. Thus, since  $W$  is continuous,

$$J(x|1-) = x + \sum_{k \neq 0} \frac{1 - e^{-i2\pi kx}}{i2\pi k} W(2\pi k).$$

But by Abel's theorem, Titchmarsh [16],  $J(x) = J(x|1-)$ . Hence finally

$$(7) \quad J(x) - x = \sum_{k \neq 0} \frac{1 - e^{-i2\pi kx}}{i2\pi k} W(2\pi k).$$

In the case  $G(x)$  is symmetric about zero, viz.  $G(x) = 1 - G(-x)$ , one has

$$(8) \quad J(x) - x = \sum_{k=1}^{\infty} W(2\pi k) \frac{\sin(2\pi kx)}{\pi k}.$$

It is now clear that the quality of the approximation is in general high and does not depend on the fine structure of  $G$  and hence of  $F$ . For only  $W$  in the neighborhood of the origin is liable to inflate the sum appreciably and this depends primarily on the nature of  $G$  at infinity.

If, for example,  $G(y)$  is Gaussian with mean zero and variance  $\sigma^2$ , then  $W(u) = \exp(-\frac{1}{2}\sigma^2 u^2)$ , and it is very clear that as  $\sigma$  increases and the tails lift the approximation improves markedly. This is in excellent accord with one's intuition.

An explicit bound on  $J(x) - x$  may be obtained by noticing that  $|1 - e^{-i2\pi kx}| \leq 2$ , and hence

$$|J(x) - x| \leq \sum_{k \neq 0} \frac{1}{\pi k} |W(2\pi k)|.$$

If  $G$  has a density  $g$ , then

$$W(2\pi k) = \frac{1}{2\pi ki} \int_{-\infty}^{\infty} e^{i2\pi kx} dg.$$

Thus

$$|W(2\pi k)| \leq \frac{1}{2\pi|k|} \int_{-\infty}^{\infty} |dg| = \frac{1}{2\pi|k|} V[g],$$

where  $V[g]$  is the variation of  $g$  on  $(-\infty, \infty)$ . Whence

$$|J(x) - x| \leq \frac{V[g]}{2\pi^2} \sum_{k \neq 0} \frac{1}{k^2} = \frac{1}{6} V[g].$$

We summarize in the following

THEOREM 2. If

$$1. \quad W(u) = \int_{-\infty}^{\infty} e^{iux} dG(x);$$

$$2. \quad W(u) = O(|u|^{-h}), \quad h > 0, |u| \rightarrow \infty;$$

then

$$\sum_{m=-\infty}^{\infty} [G(x+m) - G(m)] = \sum_{k=-\infty}^{\infty} \frac{1 - e^{-i2\pi kx}}{i2\pi k} W(2\pi k).$$

COROLLARY. If

$$1. \quad W(u) = \int_{-\infty}^{\infty} e^{iux} g(x) dx;$$

$$2. \quad V[g] < \infty;$$

$$3. \quad J(x) = \sum_{m=-\infty}^{\infty} [G(x+m) - G(m)];$$

then

$$|J(x) - x| \leq \frac{1}{6} V[g].$$

**3. Remarks.** In inventory problems one is often concerned with non-negative random variables  $X_i$ ,  $i = 1, 2, \dots$  which are independent, identically distributed, and possess a mean much smaller than some number  $K$ . One is interested in the first time  $S_n = X_1 + \dots + X_n$ ,  $n = 1, 2, \dots$ , exceeds  $K$ . Let this time be a random variable  $T$ . If the time axis is split up into contiguous intervals (periods) of length  $P$ , much smaller than the mean and variance of  $T$ , then it is often assumed that the time during the period at which the first exceedance occurs has an approximately uniform distribution. Time is here being measured from the beginning of the period. This is intuitively very appealing. Suppose  $T$  has cumulative  $G(t)$  and that the problem is scaled such that  $P = 1$ . Then intuition says  $J(x) - x$  is "small", where

$$J(x) = \sum_{n=0}^{\infty} [G(x+n) - G(n)].$$

Previous results make it clear why this is in fact so.

A close connexion exists between  $J(x) - x$  being small and Poincaré's observation on finely divided roulette wheels. Suppose the disc mentioned in the introduction is divided up into  $2n$  contiguous intervals alternately of length  $\rho$  and  $\beta$ . Let  $\rho/(\rho + \beta)$  and  $\beta/(\rho + \beta)$  be independent of  $n$ . The segments of length  $\rho$  are called red, the others black. Fréchet [5] shows, for arbitrary distributions of  $\theta$ , that the probability of obtaining red approaches  $\rho/(\rho + \beta)$  as  $n$  tends to infinity and thus similarly for black. Here the quality of the approximation is improved by shrinking the fundamental unit relative to the variance of the underlying distribution rather than increasing the variance relative to the fundamental unit.

These considerations have an obvious import for the generation of pseudo-random numbers both by electronic computers and by special purpose machines.

The foregoing results bear on questions of round-off in computing machines. Since  $d(uv) = u dv + v du$  the error resulting from multiplying two rounded numbers will be governed primarily by the first significant digits of the two numbers being multiplied. Now the distribution of first significant digits, favoring as it does low order digits, tends to produce less error than would be the case if first significant digits were uniform as has sometimes been assumed.

**Acknowledgments.** F. Mosteller and W. Kruskal provided the references to previously published material, and I am most grateful. R. Hamming was first to call my attention to the logarithmic law. He also was first to suggest that only the logarithm would satisfy the invariance principle. To one who invariably provides delightful ideas I am thankful indeed.

#### REFERENCES

- [1] J. ATICHISON, "A statistical theory of remnants," *J. Roy. Stat. Soc.*, Vol. 21 (1959), pp. 158-168.
- [2] FRANK BENFORD, "The law of anomalous numbers," *Proc. Amer. Philos. Soc.*, Vol. 78, No. 4 (1938), pp. 551-572.
- [3] G. W. BROWN, "History of RAND's random digits," *Nat. Bur. Stds., App. Math. Series*, Vol. 12 (1951), pp. 31-32.
- [4] A. DVORETSKY AND J. WOLFOWITZ, "Sums of random integers reduced modulo  $m$ ," *Duke Math. J.*, Vol. 18 (1951), pp. 501-507.
- [5] MAURICE FRÉCHET, *Calcul des Probabilités*, Vol. 2, Gauthier-Villars, Paris, 1950.
- [6] L. V. FURLAN, *Das Harmoniegesetz der Statistik*, Basel, 1946.
- [7] W. H. FURRY AND H. HURWITZ, "Distribution of numbers and distribution of significant figures," *Nature*, Vol. 155 (1945), pp. 52-53.
- [8] CORRADO GINI, "Sulla frequenza delle cifre iniziali dei numeri osservati," *Bull. Inst. Internat. Stat.*, 29th session, Vol. 35, 2<sup>a</sup> me Livraison, Rio De Janeiro (1957), pp. 57-76.
- [9] S. A. GOUDSMIT AND W. H. FURRY, "Significant figures of numbers in statistical tables," *Nature*, Vol. 154 (1944), pp. 800-801.
- [10] G. H. HARDY AND E. M. WRIGHT, *An Introduction to the Theory of Numbers*, Oxford Press, Oxford, 1954.
- [11] AMATO HERZEL, "Sulla distribuzione delle cifre iniziali dei numeri statistici," *Atti della XV e XVI Riunione, Società Italiana di Statistica*, Rome, 1957.
- [12] H. BURKE HORTON AND R. TYNES SMITH III, "A direct method for producing random digits in any number system," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 82-90.
- [13] M. KAC, *Independence in Probability: Analysis and Theory of Numbers*, John Wiley and Sons, New York, 1959.
- [14] P. LÉVY, "L'addition des variables aléatoires définies sur une circonférence," *Bull. Soc. Math. France*, Vol. 67 (1939), pp. 1-41.
- [15] HERBERT ROBBINS, "On the equidistribution of sums of independent random variables," *Proc. Amer. Math. Soc.*, Vol. 4 (1953), pp. 786-799.
- [16] E. C. TITCHMARSH, *The Theory of Functions*, Oxford University Press, Oxford, 1932.
- [17] K. D. TOCHER, "The application of automatic computers to sampling experiments," *J. Roy. Stat. Soc., Ser. B*, Vol. 16 (1954), pp. 39-61.
- [18] W. ALLEN WALLIS AND HARRY V. ROBERTS, *Statistics: A New Approach*, The Free Press, Glencoe, Illinois, 1957.

## MARKOV RENEWAL PROCESSES: DEFINITIONS AND PRELIMINARY PROPERTIES<sup>1</sup>

BY RONALD PYKE<sup>2</sup>

*University of Washington*

**1. Summary.** This paper contains the definition of and some preliminary results on Markov Renewal processes and Semi-Markov processes. The close relationship between these two types of processes is described. The concept of regularity is introduced and characterized. A classification of the states of a Markov Renewal process is described and studied.

**2. Introduction.** At the International Congress of Mathematicians held at Amsterdam in 1954, Lévy [1] and Smith [2] independently presented papers in which a new class of stochastic processes, called Semi-Markov processes (S.-M.P.) by both authors, was defined. These processes are generalizations of both continuous and discrete parameter Markov processes with countable state spaces. In the case of Lévy, the suggestion of this possible generalization is credited to K. L. Chung. Also in 1954, Takács [3] introduced essentially the same type of stochastic process, and applied them to some problems in Counter theory.

A rough, yet descriptive, definition of an S.-M.P. would be that it is a stochastic process which moves from one to another of a countable number of states with the successive states visited forming a Markov chain, and that the process stays in a given state a random length of time, the distribution function (d.f.) of which may depend on this state as well as on the one to be visited next. It is thus a Markov Chain for which the time scale has been randomly transformed.

The family of stochastic processes to be defined and studied in this paper, called Markov Renewal processes (M.R.P.), may be shown to be equivalent to the family of S.-M.P.'s. An M.R.P. is one which records at each time  $t$  the number of times a particle has visited each of the possible states up to time  $t$ , if the particle moves from state to state according to a Markov Chain and if the time required for each successive move is a random variable (r.v.) whose d.f. may depend on the two states between which the move is being made.

It will be seen, after the definition of an M.R.P. has been formalized in Section 3 below, that a Renewal process (i.e., a sequence of independent, identically distributed nonnegative r.v.'s) is equivalent to the special case of an M.R.P. with one state. However, as will become evident in the discussions below and in [4], the relationship between Renewal theory and that of M.R.P.'s is very much

---

Received January 18, 1960; revised May 9, 1961.

<sup>1</sup> The results of this paper were presented in an invited paper to the Annual Meeting of the Institute of Mathematical Statistics, December 1959.

<sup>2</sup> This research was supported by the Office of Naval Research under Contract Number Nonr-266(59), Project Number 042-205, Columbia University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

stronger than this fact alone indicates. Indeed, it would not be overexaggerating to describe the present theory as a marriage of the theories of Markov Chains and of Renewal processes. It is this close relationship which suggested the nomenclature, Markov Renewal process.

In Section 3, M.R.P.'s and S.-M.P.'s are defined, as well as some related processes. Although the processes studied here have been given "constructive" definitions, and hence are automatically separable, and have no instantaneous states (as may the S.-M.P.'s defined by Lévy [1]), there still exists the problem of whether or not an infinite number of transitions may be made in a finite interval of time. This problem is studied in Section 4, where a complete characterization is given of those M.R.P.'s for which only a finite number of transitions may be made in a finite interval of time. Such an M.R.P., with only a slight qualification, is said to be regular. It is proved that every M.R.P. with only finitely many states is regular. Several sufficient conditions for regularity are also given. In Section 5, an extension is made of the terminology used in classifying the states of a Markov Chain, to cover the case of M.R.P.'s. It is shown that the classification of any particular state in an M.R.P. is very greatly dependent upon the classification of this state in an embedded Markov Chain.

Many papers have been written on S.-M.P.'s and M.R.P.'s since 1954, mostly in the past two years. All papers known to this author which concern these processes and which are not referred to in the body of this paper are included in the supplementary references at the end of this paper, thus providing the reader with a complete list of references on this subject.

**3. Definitions and notations.** Bold face letters such as  $F$ ,  $Q$ ,  $H$ ,  $f$ ,  $q$ ,  $h$ , are consistently used in this paper to denote (real) matrix-valued functions, with the capital letters having domain  $(-\infty, \infty)$  and the lower-case letters having domain  $(0, \infty)$ . Mass functions (i.e., distribution functions whose total variations need not be equal to one) will be denoted by capital italic letters, whereas the corresponding lower-case letters will denote their respective Laplace-Stieltjes (L.-S.) transforms. For example, for  $s \geq 0$ ,  $f(s) = \int_{-\infty}^{\infty} e^{-sx} dF(x)$ , which may for the present be infinite. It will be convenient to introduce the degenerate d.f.'s,  $U_c(x) = 1$  or  $0$ , according as  $x \geq$  or  $< c$ . Unless otherwise stated, the subscripts  $i, j$  in a matrix  $(b_{ij})$  or elsewhere will run through the integers greater than or equal to 1 and not greater than  $m$ , where  $m$ , fixed, is either a finite positive integer or plus infinity. The following convolution notation is used in this and subsequent papers.  $K(t) * L(t) = \int_0^t K(t-y) dL(y)$  if  $t \geq 0$  and  $= 0$  if  $t < 0$  for functions  $K$  and  $L$  for which the Lebesgue-Stieltjes integral is defined. Write  $K * L$  for the function  $K(\cdot) * L(\cdot)$ ,  $K^{(0)} = U_0(\cdot)$ ,  $K^{(n)} = K^{(n-1)} * K$ , ( $n = 1, 2, \dots$ ) and  $K^{(-1)} = \sum_{n=0}^{\infty} K^{(n)}$  whenever the series converges.

**DEFINITION 3.1.** Let  $Q = (Q_{ij})$  be a matrix-valued function on  $(-\infty, \infty)$ .  $Q$  is called a *matrix of transition distributions* if the  $Q_{ij}$  are mass functions satisfying (i)  $Q_{ij}(t) = 0$  for  $t \leq 0$  and (ii)  $\sum_{j=1}^m Q_{ij}(+\infty) = 1$ , ( $1 \leq i < m+1$ ).

For each  $i$  and every real  $t$ , set  $H_i(t) = \sum_{j=1}^m Q_{ij}(t)$ . With this notation, (ii) of Definition 3.1 is equivalent to stating that every  $H_i$  is a d.f.



DEFINITION 3.2. The  $m \times 1$  vector  $\mathbf{A} = (a_1, a_2, \dots, a_j, \dots)$ , is called a *vector of initial probabilities* if it satisfies (i)  $a_j \geq 0$  and (ii)  $\sum_{i=1}^m a_i = 1$ .

DEFINITION 3.3. The  $(J, X)$ -process<sup>3</sup> is defined as any two-dimensional stochastic process  $\{(J_n, X_n); n \geq 0\}$  defined on a complete probability space  $(\Omega, \mathcal{B}, P)$ , that satisfies  $X_0 = 0$  a.s.,

$$(3.1) \quad P[J_0 = k] = a_k$$

and

$$(3.2) \quad P[J_n = k, X_n \leq x | J_0, J_1, X_1, J_2, X_2, \dots, J_{n-1}, X_{n-1}] \stackrel{\text{a.s.}}{=} Q_{J_{n-1}, k}(x)$$

for all  $x \in (-\infty, \infty)$  and  $1 \leq k < m + 1$ .

Set  $S_n = \sum_{i=0}^n X_i$  for  $n \geq 0$ . The  $(J, X)$ -process defined above is closely related to a Markov process as shown in

LEMMA 3.1. *The two-dimensional  $(J, S)$ -process is a Markov process, and the  $J$ -process is a Markov Chain. In particular for  $1 \leq k < m + 1, n > 0$*

$$(3.3) \quad P[J_n = k, S_n \leq y | J_0, J_1, S_1, \dots, J_{n-1}, S_{n-1}] \stackrel{\text{a.s.}}{=} Q_{J_{n-1}, k}(y - S_{n-1})$$

and

$$(3.4) \quad P[J_n = k | J_0, J_1, \dots, J_{n-1}] \stackrel{\text{a.s.}}{=} Q_{J_{n-1}, k}(+\infty).$$

PROOF. That the  $J$ -process is a Markov Chain satisfying (3.4) is an immediate consequence of (3.1), (ii) of Definition 3.1 and the Lebesgue monotone convergence theorem applied to (3.2) when  $x \rightarrow +\infty$ . That the  $(J, S)$ -process is a Markov process is implied by (3.3) and (3.1). To verify (3.3), write the left-hand side of this expression as

$$P[J_n = k, X_n \leq y - S_{n-1} | J_0, J_1, S_1, \dots, J_{n-1}, S_{n-1}].$$

Since the conditioning  $\sigma$ -field of this conditional expectation is equal to that of the left-hand side of (3.2), and since this  $\sigma$ -field is generated by a finite number of r.v.'s it is known that a conditional probability distribution (as defined by Doob [5], p. 26) exists, by means of which it is easily seen that (3.2) implies (3.3).

Because of (3.4), it is natural to define  $p_{ij} = Q_{ij}(+\infty)$  and  $\mathbf{P} = (p_{ij})$ . By Definition 3.1,  $\mathbf{P}$  is a stochastic matrix. Furthermore, if  $p_{ij} > 0$ , define  $F_{ij} = p_{ij}^{-1}Q_{ij}$ , while if  $p_{ij} = 0$  set  $F_{ij} = U_1$ . (Actually, when  $p_{ij} = 0$ ,  $F_{ij}$  may be chosen arbitrarily. There is some notational advantage, however, in choosing a d.f. which has all moments finite, but the particular choice of a degenerate d.f. has no special merit.) Set  $\mathbf{F} = (F_{ij})$ . For convenience define  $J_\infty = \infty$ . Furthermore, introduce the following notation for moments.

$$(3.5) \quad \begin{aligned} b_{ij} &= \int_0^\infty t dF_{ij}(t), & \eta_i &= \int_0^\infty t dH_i(t) \\ \sigma_{ij}^2 &= \int_0^\infty (t - b_{ij})^2 dF_{ij}(t), & \sigma_i^2 &= \int_0^\infty (t - \eta_i)^2 dH_i(t). \end{aligned}$$

<sup>3</sup> As a convenient abbreviated notation, stochastic processes will be denoted by the letter(s) used to designate the corresponding r.v.'s.

The following easily verified consequences of the above definitions will be useful in later discussions.

$$P[X_n \leq x \mid J_0, \dots, J_{n-1}] = H_{J_{n-1}}(x),$$

$$P[J_n = j \mid J_0, \dots, J_{n-1}] = p_{J_{n-1}, j},$$

$$P[X_n \leq x \mid J_0, \dots, J_n] = F_{J_{n-1}, J_n}(x),$$

$$\begin{aligned} (3.6) \quad P[X_{n_1} \leq x_1, X_{n_2} \leq x_2, \dots, X_{n_k} \leq x_k \mid J_n; n \geq 0] \\ = P[X_{n_1} \leq x_1, \dots, X_{n_k} \leq x_k \mid J_0, J_1, \dots, J_{n_k}] \\ = \prod_{i=1}^k F_{J_{n_i-1}, J_{n_i}}(x_i) \end{aligned}$$

for  $0 < n_1 < \dots < n_k$ , all equalities holding with probability one. It follows from the last two relationships that  $X_{n_1}, X_{n_2}, \dots, X_{n_k}$  are mutually conditionally independent given  $J_{n_1-1}, J_{n_1}, \dots, J_{n_k-1}, J_{n_k}$  (e.g., cf. [6], Definition 3).

In Renewal theory, the basic process studied is that which gives the number of partial sums or renewals in the intervals  $(0, t]$  for all  $t \geq 0$ . The natural analogues to this for the present theory are the counting processes defined now.

**DEFINITION 3.4.** The integer-valued stochastic processes  $\{N(t); t \geq 0\}$  and  $\{N_j(t); t \geq 0\}$  are defined by  $N(t) = \sup \{n \geq 0: S_n \leq t\}$  and  $N_j(t) = \text{no. of times } J_k = j \text{ for } 0 < k \leq N(t) + 1$ .

Notice that without added restrictions on  $m$  and/or  $\mathbf{Q}$ ,  $N(t)$  may be infinite with positive probability. Notice also that the counting functions  $N_j$  are defined so as not to record the value of  $J_0$ . Setting  $\mathbf{N}(t) = (N_1(t), N_2(t), \dots, N_j(t), \dots)$ , the stochastic process  $\{\mathbf{N}(t); t \geq 0\}$  is called a *Markov Renewal Process* (M.R.P.) determined by  $(m, \mathbf{A}, \mathbf{Q})$ . Clearly  $N(t) = \sum_{j=1}^m N_j(t)$  a.s.

Related to an M.R.P. is the stochastic process defined now which simply records the state of the process at each time point.

**DEFINITION 3.5.** The  $Z$ -process,  $\{Z_t; t \geq 0\}$  defined by  $Z_t = J_{N(t)}$  is called a *Semi-Markov Process* (S.-M.P.) determined by  $(m, \mathbf{A}, \mathbf{Q})$ .

Let us introduce some additional vocabulary to facilitate later discussions. We shall say that a "transition" of an M.R.P. has occurred at each of the time points  $S_0, S_1, S_2, \dots$ . The process (either an M.R.P. or an S.-M.P.) is said to be "in state  $i$ " at time  $t$ , if, and only if,  $Z_t = i$ .

As defined in Definition 3.4, an M.R.P. is a vector-valued process (infinite dimensional if  $m = \infty$ ). It is clear that one could construct one-dimensional processes that are probabilistically equivalent to the  $\mathbf{N}$ -process. For example, the  $Y$ -process defined by  $Y_t = j + 1 - 2^{-k+1}$  on the set  $[J_{N(t)-n} = j, 0 \leq n < k, J_{N(t)-k} \neq j]$  and  $= \infty$  on the set  $[N(t) = \infty]$  may be shown to be equivalent to the  $\mathbf{N}$ -process, since it records both the state of the M.R.P. and the number of preceding consecutive transitions to state  $i$  for each  $t > 0$ . For most discussions, the r.v.'s  $N_j(t)$ , and especially their expectations, play the central role, as does  $N(t)$  for the special case of a Renewal process, namely the case  $m = 1$ . However,

the  $Y$ -process representation serves to emphasize the relationships between an M.R.P. and an S.-M.P. The  $Y$ -process is always an S.-M.P., and is called the *associated S.-M.P.* of the given M.R.P. It is equal almost surely to the  $Z$ -process if and only if,  $p_{ii} = 0$  for every state  $i$  which can be reached with positive probability. Otherwise the  $Y$ -process always has an infinity of states regardless of the value of  $m$ . It follows from this  $Y$ -process representation of an M.R.P. that, theoretically at least, any results about an M.R.P. may be derived from theorems concerning S.-M.P.'s. It is, however, both convenient and practical to keep these two kinds of processes distinct, and to use the process most natural for a given problem. For computation of moments of recurrence times (cf. [7]), it is natural to work this out for M.R.P.'s since most applications involve processes in which a transition from a state to itself is possible. It should be observed that although an M.R.P. has a finite number of states, the associated S.-M.P. will in most applications have an infinite number of states, as is the case, for example, for a Renewal process ( $m = 1$ ). On the other hand, for problems concerning the limiting stationarity of a given M.R.P., transitions from a state to itself play no essential role. One may then, without loss of generality, work with the related matrix of transition distributions  $Q^* = (Q_{ij}^*)$  defined by  $Q_{ii}^* = Q_{ii}$  if  $p_{ii} = 1$ ,  $Q_{ii}^* = 0$  if  $p_{ii} < 1$  and  $Q_{ij}^* = Q_{ij}[1 - Q_{ii}]^{(-1)}$  if  $i \neq j$  and  $p_{ii} < 1$ . One may verify

LEMMA 2. *Every S.-M.P. determined by  $(m, A, Q)$  has the same family of joint d.f.'s as every S.-M.P. determined by  $(m, A, Q^*)$ .*

Any S.-M.P. determined by  $(m, A, Q^*)$  is called a *corresponding S.-M.P.* of the given M.R.P.

When  $m = 1$ , a Markov Renewal process becomes a Renewal process, the theory of which is extensive (cf. the survey paper on Renewal theory by Smith [8]). When the transition distributions are of the form  $Q_{ij} = p_{ij}U_1$  for all  $i$  and  $j$ , the Markov Renewal process becomes a Markov Chain, and in this case is equivalent to its corresponding S.-M.P. by virtue of the constant transition times. Moreover, a continuous parameter Markov process with  $m$  states, *all of which are stable*, is a special case of an M.R.P. (in fact, of an S.-M.P.) for which the  $Q_{ij}$  are of the form

$$(3.7) \quad Q_{ij}(t) = p_{ij} \max(0, 1 - e^{-\lambda_i t}) \quad (-\infty < t < \infty)$$

for constants  $\lambda_i > 0$ , and  $p_{ii} = 0$  for every  $i$ .

**4. Finiteness of  $N(t)$  and regularity.** It may easily be deduced from the constructive definitions of an M.R.P. and an S.-M.P. given in Section 3 that they are separable and that almost all sample functions of the  $Y$ -process, and of the  $Z$ -process, are step-functions over an interval of the form  $[0, L)$  and identically equal to infinity over  $[L, \infty)$ , where  $L > 0$  is a possibly infinite r.v. which is a Borel function of the  $Y$ -process. Clearly, the sets  $[L < \infty]$  and  $[N(t) = \infty, t \geq L]$  differ only by a set of measure zero. It is important to be able to characterize those M.R.P.'s for which  $L \stackrel{a.s.}{=} \infty$ , or equivalently, those for which

$N(t) < \infty$  for all  $t$ . We shall first verify the intuitive result that in the case of  $m < \infty$ , the (a.s.) finiteness of  $N(t)$  is always true.

LEMMA 4.1. *If  $m < \infty$ , then for all states  $i$ ,*

$$(4.1) \quad P[N(t) < \infty, \text{ for all } t \geq 0] = 1.$$

PROOF. By Definition 3.4,  $N(t)$  is nondecreasing. It suffices, therefore, to prove that  $P[N(t) < \infty | Z_0 = i] = 1$  for each  $t \geq 0$ , and for every  $i$  for which  $a_i > 0$ . Suppose  $a_i > 0$ . By Definition 3.4, (3.3) and (3.4) one obtains for  $t \geq 0$

$$\begin{aligned} P[N(t) \geq n | Z_0 = i] &= P[S_n \leq t | J_0 = i] \\ &= \sum_{s_n, i} \sum_{j=0}^{n-1} Q_{\alpha_j \alpha_{j+1}}(t) \\ &= \sum_{s_n, i} \left\{ \prod_{j=0}^{n-1} p_{\alpha_j \alpha_{j+1}} \right\} \sum_{j=0}^{n-1} F_{\alpha_j \alpha_{j+1}}(t) \end{aligned}$$

where  $*$  denotes convolution of the indicated d.f.'s and where

$$(4.2) \quad S_{n,i} = \{(\alpha_0, \alpha_1, \dots, \alpha_n) : \alpha_0 = i, \alpha_j \text{ an integer}, 1 \leq \alpha_j \leq m(1 \leq j \leq n)\}$$

is the set of all paths of length  $n+1$  of the  $J$ -process for which  $J_0 = i$ . Define  $F = \max_{i,j} F_{ij}$ . It is well known that for mass functions  $F_1, F_2, G_1$  and  $G_2$  for which  $F_1 \leq G_1$  and  $F_2 \leq G_2$ , one has  $F_1 * F_2 \leq G_1 * G_2$ . Consequently, it follows from (3.6) that

$$P[N(t) \geq n | Z_0 = i] \leq F^{(n)}(t) \sum_{s_n, i} \prod_{j=0}^{n-1} p_{\alpha_j \alpha_{j+1}} = F^{(n)}(t).$$

Since  $m < \infty$ , one has by Definition 3.1 that  $F(0) = 0$  and so for  $t > 0$ ,  $F^{(n)}(t) \rightarrow 1$  as  $n \rightarrow +\infty$ .

Any d.f.  $F$  satisfying  $F(0) = 0$  which is an upper bound for every  $F_{ij}$ , would have sufficed in the proof of Lemma 4.1. An alternative choice of  $F$  which has a more intuitive interpretation than that used in the above proof is

$$F = 1 - \prod_{i,j} [1 - F_{ij}],$$

the d.f. of the minimum of a family of independent r.v.'s, one corresponding to each d.f.  $F_{ij}$ .

A consequence of Lemma 4.1 is that almost all path functions of a  $Y$ -process with  $m < \infty$  are step-functions over  $[0, \infty)$ , as is also true for the corresponding S.-M.P.

Consider now the case of unrestricted  $m$ . For this case, it is necessary to impose restrictions in order to insure the (a.s.) finiteness of  $N(t)$ . To see this, the simplest example is the degenerate one for which  $Q_{j,j+1} = U_{2^{-j}}(\cdot) (j \geq 1)$  and all other  $Q_{ij} = 0$ . For this example,  $N(t) \stackrel{\text{a.s.}}{=} n$ , whenever  $1 - 2^{-n} \leq t < 1 - 2^{-(n-1)}$  for  $n \geq 0$ , while  $N(t) \stackrel{\text{a.s.}}{=} \infty$ , whenever  $t \geq 1$ . That is,  $L = 1$  (a.s.). In

what follows, a necessary and sufficient condition for the (a.s.) finiteness of  $N(t)$  for every  $t \geq 0$  is given, as well as several sufficient conditions which are applicable in the more common situations.

For any  $c > 0$ , define the truncated moments  $b_{ij}^{(c)} = \int_0^c t dF_{ij}(t)$ . Clearly  $b_{ij} = \lim_{c \rightarrow \infty} b_{ij}^{(c)}$ . Define the family of integer sequences

$$S_i = \{(\alpha_0, \alpha_1, \dots) : \alpha_0 = i, 1 \leq \alpha_j < m + 1 (j \geq 1)\}.$$

DEFINITION 4.1. A state  $i$  of an M.R.P. determined by  $(m, \mathbf{A}, \mathbf{Q})$  is said to be *regular-A* if either  $P[(J_0, J_1, \dots) \in S_i] = a_i = 0$ , or  $a_i > 0$  and there exists a measurable subset  $\alpha \subset S_i$  such that  $P[(J_0, J_1, \dots) \in \alpha | J_0 = i] = 1$  and such that for every  $(\alpha_0, \alpha_1, \dots) \in \alpha$  and for every  $c > 0$  at least one of the series

$$(4.3) \quad \sum_{j=0}^{\infty} [1 - F_{\alpha_j \alpha_{j+1}}(c)], \quad \sum_{j=0}^{\infty} b_{\alpha_j \alpha_{j+1}}^{(c)}$$

diverges. An M.R.P. determined by  $(m, \mathbf{A}, \mathbf{Q})$  is said to be *regular-A* if each of its states is *regular-A*. If these properties hold for all initial distributions  $\mathbf{A}$ , the state or the M.R.P. will be called *regular*. Since whether an M.R.P. is regular or not depends only upon the nature of  $\mathbf{Q}$ , we shall alternatively speak of  $\mathbf{Q}$  as being regular.

It would have sufficed in the above definition to have required the divergence of one of the series in (4.3) for only those sequences in  $\alpha$  for which  $p_{\alpha_j \alpha_{j+1}} > 0$  for every  $j$ . This is so because of the convention made earlier, that whenever  $p_{ij} = 0$ ,  $F_{ij} = U_1$  and hence  $b_{ij}^{(c)} = 1 > 0$  for all  $c \geq 1$ . It is shown in the following theorems that the concepts of the above definition may be used to characterize the (a.s.) finiteness of  $N(t)$ .

THEOREM 4.1. For any given state  $i$  of an M.R.P. determined by  $(m, \mathbf{A}, \mathbf{Q})$ ,

$$(4.4) \quad P[J_0 = i, N(t) = \infty \text{ for some } t \geq 0] = 0$$

if and only if  $i$  is *regular-A*.

PROOF. The theorem is obvious whenever  $a_i = 0$ . Assume, therefore, that  $a_i > 0$ . For any  $(\alpha_0, \alpha_1, \dots) \in S_i$ , one can show that

$$(4.5) \quad \begin{aligned} P[N(t) = \infty | J_k = \alpha_k, k \geq 0] &= \lim_{n \rightarrow \infty} P[S_n \leq t | J_k = \alpha_k, k \geq 0] \\ &= \sum_{j=0}^{\infty} F_{\alpha_j \alpha_{j+1}}(t) \end{aligned}$$

by the last relationship of (3.6). It is known, and easily verified, that for non-negative r.v.'s Kolmogorov's Three-Series criterion (cf. [9], p. 236) for a.s. convergence of a series of independent r.v.'s, becomes a "two-series" criterion, namely, "If  $\{V_n : n \geq 1\}$  is a sequence of nonnegative independent r.v.'s, then the series  $\sum_{n=0}^{\infty} V_n < \infty$  (a.s.) if, and only if, for some finite  $c > 0$ ,

$$\sum_{n=1}^{\infty} P[V_n > c] < \infty, \quad \text{and} \quad \sum_{n=1}^{\infty} E[\min(V_n, c)] < \infty.$$

Furthermore, if the series does not converge a.s. then it diverges a.s." Now (4.5) implies that with respect to the indicated conditional probability measure, the  $S_n$ 's are representable as partial sums of independent nonnegative r.v.'s. Therefore, by the above version of Kolmogorov's theorem, one has that

$$(4.6) \quad \sum_{j=0}^{\infty} {}^* F_{\alpha_j \alpha_{j+1}}(t) = 0 \quad (t \geq 0)$$

if, and only if, for all  $c > 0$ , at least one of the series

$$\sum_{n=1}^{\infty} P[X_n > c \mid J_k = \alpha_k, k \geq 0], \quad \sum_{n=1}^{\infty} E[\min(X_n, c) \mid J_k = \alpha_k, k \geq 1]$$

diverges, which is easily checked to be equivalent to specifying that at least one of the series given in (4.3) diverges. Therefore, if state  $i$  ( $\alpha_i > 0$ ) is regular-A, there exists a set  $\mathcal{Q} \subset \mathcal{S}_i$  of conditional probability equal to one, such that for every  $(\alpha_0, \alpha_1, \dots) \in \mathcal{Q}$ , (4.6) is satisfied, and hence by (4.5)

$$P[N(t) = \infty \mid J_0 = i] = 0,$$

thus verifying (4.4). Conversely, if (4.4) is satisfied, then

$$0 = P[N(t) = \infty \mid J_0 = i] = E \left[ \sum_{n=0}^{\infty} {}^* F_{J_n J_{n+1}}(t) \mid J_0 = i \right].$$

Because of the nonnegativeness of the integrand, this implies

$$P \left[ \sum_{n=0}^{\infty} {}^* F_{J_n J_{n+1}}(t) = 0 \mid J_0 = i \right] = 1.$$

Consequently, in Definition 4.1, one may choose  $\mathcal{Q} \subset \mathcal{S}_i$  to be the set of all  $\alpha$ -sequences satisfying (4.6). Hence, state  $i$  is regular-A.

**COROLLARY 4.1.** For an M.R.P. determined by  $(m, \mathbf{A}, \mathbf{Q})$ ,

$$P[N(t) < \infty \text{ for all } t] = 1$$

if, and only if, it is regular-A.

**COROLLARY 4.2.** For a given  $m$  and  $\mathbf{Q}$ ,

$$P[N(t) < \infty \text{ for all } t] = 1$$

for all choices of a vector of initial probabilities if, and only if,  $\mathbf{Q}$  is regular.

The foregoing theorems give a complete characterization of those M.R.P.'s having almost all sample functions equal to step-functions over  $(0, \infty)$ . In many practical situations, due to additional assumptions being stated, it is not necessary to check completely the conditions for regularity as given in Definition 4.1. In many instances, weaker sufficient conditions are available, and possibly are more easily checked. Some of these are given in the following discussion.

The simplest sufficient condition is a consequence of Lemma 4.1 namely that if  $m < \infty$ , then the M.R.P. is regular.

It is also easily shown that if for each  $\alpha$ -sequence in a subset  $\mathcal{Q} \subset \mathcal{S}_i$  of (con-



ditional) probability one, there exists a finite  $M > 0$ , such that  $F_{\alpha_j \alpha_{j+1}}(M) = 0$ , then state  $i$  is regular. A particular application of this is to Markov Chains over discrete time, of which all states must, therefore, be regular.

If  $\sum_{j=0}^{\infty} b_{\alpha_j \alpha_{j+1}} < \infty$  on some set of  $\alpha$ -sequences in  $S_i$  of positive (conditional) measure, then  $i$  is not a regular state. This is a simple consequence of the fact that the convergence of the series of expectations of a sequence of independent r.v.'s, implies the convergence (a.s.) of the series of r.v.'s.

If for every  $\alpha$ -sequence in a subset  $\mathcal{A} \subset S_i$  of (conditional) probability one, either  $\sum_{j=0}^{\infty} \sigma_{\alpha_j \alpha_{j+1}}^2 < \infty$ , or there exists a finite  $M > 0$  (possibly depending on the sequence) such that  $F_{\alpha_j \alpha_{j+1}}(M) = 1$  ( $j \geq 0$ ), then state  $i$  is regular if and only if  $\sum_{j=0}^{\infty} b_{\alpha_j \alpha_{j+1}} = \infty$  for every  $(\alpha_0, \alpha_1, \dots) \in \mathcal{A}$ . This again is a simple consequence of known results on sums of (positive) independent r.v.'s (cf. [9], p. 236).

As a special case of the above, consider the following sufficient condition. If for every  $\alpha$ -sequence in a subset  $\mathcal{A} \subset S_i$  of (conditional) probability one, either  $\sum_{j=0}^{\infty} \sigma_{\alpha_j \alpha_{j+1}}^2 < \infty$ , or there exists a finite  $M > 0$  (possibly depending on the sequence) such that  $F_{\alpha_j \alpha_{j+1}}(M) = 1$  ( $j \geq 0$ ), and there exist two real sequences  $\{\delta_j\}, \{\eta_j\}$  of positive numbers such that  $\sum_{j=0}^{\infty} \delta_j \eta_j = \infty$  and  $F_{\alpha_j \alpha_{j+1}}(\delta_j) \leq 1 - \eta_j$  ( $j \geq 0$ ) for every  $(\alpha_0, \alpha_1, \dots) \in \mathcal{A}$ , then state  $i$  is regular. This follows immediately from the preceding since under these conditions  $b_{\alpha_j \alpha_{j+1}} \geq \delta_j \eta_j$ . This condition is a corrected version of one due to Smith [2] (see also [10]). (When reading [2], the reader should note the different meaning of the word regular as it is used there.)

Consider now a condition designed primarily for continuous parameter Markov processes with an at most countable number of states. If there exists a set  $\{\lambda_{ij}: 1 \leq i, j < m+1\}$  of finite positive numbers such that for every  $\alpha$ -sequence in a subset  $\mathcal{A} \subset S_i$  of probability one,  $F_{\alpha_j \alpha_{j+1}}(t) = 1 - \exp(-\lambda_{\alpha_j \alpha_{j+1}} t)$  for all  $t \geq 0$ , then state  $i$  is regular if, and only if,

$$(4.7) \quad \sum_{j=0}^{\infty} \lambda_{\alpha_j \alpha_{j+1}}^{-1} = \infty$$

for all  $(\alpha_0, \alpha_1, \dots) \in \mathcal{A}$ . That regularity implies (4.7) is immediate. It suffices to show that (4.7) implies the divergence of one of the series in (4.3). This is best seen by simply evaluating the series (4.3) to be

$$\sum_{j=0}^{\infty} \exp(-c \lambda_{\alpha_j \alpha_{j+1}}),$$

$$\sum_{j=0}^{\infty} \{\lambda_{\alpha_j \alpha_{j+1}}^{-1} [1 - \exp(-c \lambda_{\alpha_j \alpha_{j+1}})] - c \exp(-c \lambda_{\alpha_j \alpha_{j+1}})\}.$$

Assuming (4.7), one has that if the first series converges, then for  $j$  sufficiently large  $1 - \exp(-c \lambda_{\alpha_j \alpha_{j+1}}) > \frac{1}{2}$ , and so the second series must diverge. For  $\lambda_{ij} = \lambda_i$ , one obtains the known result for Markov processes, and if, moreover,

one has  $p_{j,j+1} = 1$ , the above is the well known result for pure Birth processes (cf. [11] p. 349, [5] p. 271 and [12] p. 406).

Lastly, we mention a sufficient condition which is in terms of the underlying Markov Chain. If state  $i$ , considered as a state of the  $J$ -process, a Markov Chain by Lemma 3.1, is such that with (conditional) probability one, the  $J$ -process, starting in state  $i$ , will reach a recurrent state (cf. [12] p. 353), then state  $i$  is regular. To see this let  $j$  (possibly equal to  $i$ ) be the first recurrent state that is reached, and suppose it was reached at the  $n_0$ th transition ( $n_0$  may be zero). Let  $n_1, n_2, \dots$  be the successive integers  $n$  at which  $J_n = j$ . Set  $T_0 = S_{n_0}$  and  $T_k = S_{n_k} - S_{n_{k-1}}$ . By assumption, the r.v.  $n_k$  are finite a.s., and hence, so are the  $T_k$ 's. Furthermore,  $\{T_k : k \geq 1\}$  forms a Renewal process and so  $\sum_{k=1}^{\infty} T_k = \infty$  (a.s.). Since  $\sum_{n=1}^{\infty} X_n = \sum_{k=0}^{\infty} T_k$  (a.s.), the former series diverges and hence  $N(t) < \infty$  (a.s.) for every  $t$ , as required.

**5. Classification of states.** In this section, the states of an M.R.P. will be classified in much the same manner as is done for Markov Chains, the terminology of the latter being retained. The reader is referred to Feller [12] and to Chung [13] for material on Markov Chains. To facilitate the definitions to follow, it is assumed that for all M.R.P.'s considered below, every initial probability is positive. Consider the notation defined, for all  $i, j$  and  $t \geq 0$ , by

$$(5.1) \quad P_{ij}(t) = \begin{cases} P[Z_t = j \mid Z_0 = i] & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

$$(5.2) \quad G_{ij}(t) = \begin{cases} P[N_j(t) > 0 \mid Z_0 = i] & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

and let the moments (possibly infinite) of  $G_{ij}$  be denoted by  $\mu_{ij}$ . According to these definitions,  $P_{ij}(t)$  is the probability that an M.R.P., initially in state  $i$ , is in state  $j$  at time  $t$ , while  $G_{ij}$  is a mass function representing the probability distribution of the time (first passage time) until the next transition into state  $j$  of a process which is initially in state  $i$ . Notice that when  $i = j$ , this definition does not require the process to leave state  $i$  during the first passage time.  $\mu_{ii}$  will be called the mean recurrence time of state  $i$ .

**DEFINITION 5.1.** (a) States  $i$  and  $j$  are said to *communicate* if, and only if, either  $G_{ij}(\infty)G_{ji}(\infty) > 0$  or  $i = j$ .

(b) Communication is an equivalence relation, and the disjoint equivalence classes are called *classes* and are denoted by  $C_i$  (whenever  $i \in C_i$ ).

(c) An M.R.P. is said to be *irreducible* if, and only if, there is only one class.

(d) A class  $C_i$  is said to be *essential* (or closed) if, and only if, for all  $j \in C_i$  and for all  $t \geq 0$ ,  $\sum_{k \in C_i} P_{jk}(t) = 1$ .

(e) State  $i$  is said to be *recurrent* (or persistent (Smith [8]) or ergodic (Lévy [11])) if and only if  $G_{ii}(\infty) = 1$ , and is said to be *transient* otherwise.

(f) State  $i$  is said to be *null recurrent* (weakly ergodic (Lévy [11])) if, and only if, it is recurrent and  $\mu_{ii} = \infty$ . State  $i$  is said to be *positive* (ergodic (Feller [12]), strongly ergodic (Lévy [11])) if, and only if, it is recurrent and  $\mu_{ii} < \infty$ .

As should be expected, the properties defined here for M.R.P.'s are very closely related to those of the corresponding Markov Chains (c.M.C.) determined by the same  $m$ ,  $\mathbf{A}$ , and  $\mathbf{P}$ , namely, the  $J$ -processes. This is illustrated by the following theorem.

**THEOREM 5.1.** *For a given M.R.P.: (a) state  $i$  is recurrent (is transient) [communicates with state  $j$ ], if and only if state  $i$  is recurrent (is transient) [communicates with state  $j$ ] in the c.M.C.; (b) a class is essential if and only if in the c.M.C. it is essential; and (c) if  $m < \infty$ , then state  $i$  is positive if and only if state  $i$ , in the c.M.C., is positive and  $\eta_j < \infty$  for all  $j \in C_i$ .*

**PROOF:** (a) From (5.2), it follows that

$$(5.3) \quad G_{ij}(\infty) = P[J_n = j \text{ for some } n > 0 \mid J_0 = i].$$

This relation suffices to verify (a) since the properties of recurrence, transience and communication involve only the quantities  $G_{ij}(\infty)$ , and since (5.3) shows that these quantities are identical to the analogous quantities of the c.M.C. (b) is immediate. To prove (c), let  $S_{n,i,j}$  denote the subset of  $(\alpha_0, \dots, \alpha_n) \in S_{n,i}$  for which  $\alpha_n = j$ , where  $S_{n,i}$  is as defined in (4.2). Then one may straightforwardly show (for arbitrary  $m$ ) that

$$(5.4) \quad \mu_{ij} = \sum_{n=1}^{\infty} \sum_{S_{n,i,j}} \prod_{k=0}^{n-1} p_{a_k a_{k+1}} (b_{a_0 a_1} + \dots + b_{a_{n-1} a_n})$$

Now assuming  $m < \infty$  and  $i = j$ , one can write

$$(\min_{k, j \in C_i} b_{kj}) \mu_{ii}^* \leq \mu_{ii} \leq (\max_{k, j \in C_i} b_{kj}) \mu_{ii}^*$$

where

$$(5.5) \quad \mu_{ii}^* = \sum_{n=1}^{\infty} n \sum_{S_{n,i,i}} \prod_{k=0}^{n-1} p_{a_k a_{k+1}}$$

is the mean recurrence time of state  $i$  in the c.M.C. Since  $m < \infty$  the minimum shown above is positive and since  $m < \infty$  and  $\eta_j < \infty$  for all  $j \in C_i$ , the maximum shown above is finite.

There does not seem to be any simple necessary and sufficient condition for a positive state in the case of  $m = \infty$ . Examples may readily be constructed to show that a state of an M.R.P. may be positive (null recurrent), while the same state in the c.M.C. is null recurrent (positive). One sufficient condition for the positivity of state  $i$  is that the state be positive in the c.M.P. and  $\sum_{j \in C_i} \eta_j < \infty$ . The proof of this, as well as further discussion along these lines, is contained in [7] where explicit computations of the  $\mu_{ij}$  in terms of the  $\mu_{ij}^*$  is made.

#### REFERENCES

- [1] PAUL LÉVY, "Systèmes Semi-Markoviens à au plus une infinité dénombrable d'états possibles," *Proc. Int. Congr. Math.*, Amsterdam, Vol. 2 (1954), p. 294. "Processus Semi-Markoviens," *ibid.*, Vol. 3 (1954), pp. 416-426.
- [2] W. L. SMITH, "Regenerative stochastic processes," *Proc. Roy. Soc. (London)*, Ser. A, Vol. 232 (1955), pp. 6-31 [cf. the abstract of this paper in *Proc. Int. Congr. Math.*, Amsterdam, Vol. 2 (1954), pp. 304-305].

- [3] LAJOS TAKÁCS, "Some investigations concerning recurrent stochastic processes of a certain type," *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, Vol. 3 (1954), pp. 115-128 [Hungarian. English Summary. Cf. *Math. Reviews*, Vol. 17 (1956), p. 866.]
- [4] RONALD PYKE, "Markov Renewal processes with finitely many states," *Ann. Math. Stat.*, Vol. 32 (1961), pp. 1243-1259.
- [5] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.
- [6] M. V. JOHNS AND RONALD PYKE, "On conditional expectation and quasi-rings," *Pacific J. Math.*, Vol. 9 (1959), pp. 715-722.
- [7] RONALD PYKE, "Limit theorems for Markov Renewal processes," Technical Report No. 24 (1961), Contract No. Nonr-266(59) Project No. 042-205. Columbia University.
- [8] W. L. SMITH, "Renewal theory and its ramifications," *J. Roy. Stat. Soc., Ser. B*, Vol. 20 (1958), pp. 243-302.
- [9] M. LOÈVE, *Probability Theory*, 1st Edition, D. Van Nostrand Co., Princeton, N. J., 1955.
- [10] W. L. SMITH, "Remarks on the paper 'Regenerative stochastic processes'," *Proc. Roy. Soc. (London)*, Ser. A, Vol. 256 (1960), pp. 496-501.
- [11] PAUL LÉVY, "Systèmes Markoviens et stationnaires. Cas dénombrables," *Ann. École Norm. Suppl.*, Vol. 68 (1951), pp. 327-381.
- [12] W. FELLER, *An Introduction to Probability Theory*, 2nd Edition, John Wiley and Sons, New York, 1950.
- [13] K. L. CHUNG, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, Berlin, 1960.

## SUPPLEMENTARY REFERENCES

- [14] PHILIP M. ANSELONE, "Ergodic theory for discrete Semi-Markov chains," *Duke Math. J.*, Vol. 27 (1960), pp. 33-40.
- [15] RICHARD BARLOW, "Applications of Semi-Markov processes to counter and reliability problems," Technical Report No. 57 (1960), Stanford University, Contract N6onr-25140 (NR-342-022).
- [16] VIOLET R. CANE, "Behaviour sequences as Semi-Markov chains," *J. Roy. Stat. Soc., Ser. B*, Vol. 21 (1959), pp. 36-58.
- [17] AUGUSTUS J. FABENS, "The solution of queueing and inventory models by Semi-Markov processes," Technical Report No. 20 (1959), Stanford University, Contract Nonr-225(28), (NR-047-019).
- [18] HARRY KESTEN, "Occupation times for Markov and Semi-Markov chains," cf. Abstract No. 61T-77, *Notices Amer. Math. Soc.*, Vol. 8 (1961), p. 179.
- [19] RUPERT G. MILLER, JR., "Asymptotic multivariate occupation time distributions for Semi-Markov processes," Technical Report No. 70 (1961), Stanford University, Contract Nonr-225(52), (NR-342-022).
- [20] STEVEN OREY, "Change of time scale for Markov processes," Technical Report, University of Minnesota (1959), Contract No. AF 49(638)-617.
- [21] RONALD PYKE, "Markov Renewal processes of zero order and their application to counter theory," Technical Report No. 64 (1960), Stanford University, Contract Nonr-255(52), (NR-342-022); to appear in a volume edited by Samuel Karlin, Stanford University Press, Calif.
- [22] RONALD PYKE, "Doebelin Ratio limit theorems for Markov Renewal processes," to appear [cf., Abstract, *Ann. Math. Stat.*, Vol. 31 (1960), pp. 245-246].
- [23] GEORGE H. WEISS, "On a Semi-Markovian process with a particular application to reliability theory," NAVORD Report 4351, White Oak, Maryland.
- [24] CYRUS DERMAN, "Remark concerning two-state semi-Markov processes," *Ann. Math. Stat.*, Vol. 32 (1961), pp. 615-616.

# MARKOV RENEWAL PROCESSES WITH FINITELY MANY STATES<sup>1</sup>

BY RONALD PYKE

*University of Washington<sup>2</sup>*

**1. Summary.** In this paper, Markov Renewal processes having a finite number of states are studied. Explicit expressions are derived for the distribution functions of first passage times, as well as for the marginal distribution function of the corresponding Semi-Markov process. Double generating functions are obtained for the distribution functions of the  $N_j$ -processes. The limiting behavior of a Markov Renewal process is discussed, the stationary probabilities being derived completely. General Markov Renewal processes are introduced, and a related stationary process is determined. Several examples are given.

**2. Introduction.** In [1], a class of stochastic processes, called Markov Renewal processes (M.R.P.), are defined and a preliminary investigation is made of their structure, and of the related Semi-Markov processes (S.-M.P.) introduced by Lévy [2], Smith [3] and Takács [4] independently in 1954. The reader is referred to [1] for the necessary definitions and notation. Roughly speaking, M.R.P.'s are generalizations both of continuous and discrete parameter Markov Chains which permit arbitrary distribution functions (d.f.), possibly depending both on the last state entered and on the next state to be entered, for the times between successive transitions.

In the present paper we restrict our attention to those M.R.P.'s determined by  $(m, \mathbf{A}, \mathbf{Q})$  with  $m < \infty$ . Recall, that because of Lemma I.4.1,<sup>3</sup> all such M.R.P.'s are regular, i.e., almost all sample functions are finite-valued step functions over  $(-\infty, \infty)$ . In Section 3, systems of integral equations are given for the functions  $P_{ij}(t)$  and  $G_{ij}(t)$ , respectively, the d.f. of  $Z_t$  and the d.f. of the time until the first transition into state  $j$ , both given  $Z_0 = i$ . Equations relating these two functions are also given. Lemma 3.2 is due to Takács (equations (10) and (11) of [4]),<sup>4</sup> while (3.2), summed over  $i$  with respect to the initial probabilities, is essentially equation (8) in a paper by Weiss [5], who derived this and other integral relationships for the purpose of studying the asymptotic behavior of

Received January 18, 1960; revised May 26, 1961.

<sup>1</sup> This research was supported by the Office of Naval Research under Contract Number Nonr-266(59), Project Number 042-205. Reproduction in whole or in part is permitted for any purpose of the United States Government.

<sup>2</sup> This work was completed while the author was at Columbia University. The contents of the first 6 sections comprised an Invited paper given at the Annual meeting of the I.M.S., Sept. 1958, Cambridge, Massachusetts.

<sup>3</sup> All numbers which are prefixed by I, refer to the correspondingly numbered part of [1].

<sup>4</sup> In assumption 4 of [4], which postulates that the  $X$ -process is a sequence of independent random variables, the word independent should be interpreted as conditional independence given the successive states of the system.



the "Renewal functions"  $M_{ij}(t)$ , defined in (5.10). The relationships of Section 3 are solved in Section 4, with explicit expressions being given, respectively, in Theorems 4.1 and 4.2 for the matrix-valued functions  $\mathcal{O}$  and  $\mathcal{G}$  defined therein. In Section 5, the marginal d.f.'s for the  $N_j$ -processes are studied, explicit expressions being given for their double generating functions. As a consequence, the matrix-valued Renewal function  $\mathfrak{M}$  is derived. The three most common subclasses of M.R.P.'s are briefly discussed in the following section; they are Markov Chains, continuous parameter Markov processes, and Renewal processes.

In Section 7, the problem of characterizing the limiting behavior of an M.R.P. is solved. The stationary distribution of the process is derived, thus extending a result of Smith [3]. The concept of a general Markov Renewal process (G.M.R.P.) is introduced, in which a different matrix of transition distributions may be used to determine the initial transition of the process then is used for all remaining transitions. By using the stationary probabilities obtained for a given M.R.P. it is possible to construct a related G.M.R.P. which is "stationary" in a sense made explicit in Theorem 7.2. Following this, some additional examples are briefly stated in the final section.

**3. The probabilities  $P_{ij}$  and  $G_{ij}$  and related quantities.** Of considerable importance in studying the behavior of M.R.P.'s are the times between successive occurrences of a given state. It is the purpose of this section to study the d.f.'s of these times as well as the marginal d.f.'s of the corresponding S.-M.P., the  $Z$ -process, which is to say, the probability of the process being in a given state at any instant of time. Several identities and relationships between these quantities will be given for them in terms of the basic  $\mathcal{Q}$ -matrix.

Let us assume that  $\mathbf{A} > 0$  coordinate-wise for every M.R.P. under consideration, in order that all conditional probabilities will be well defined. The quantities in question are then defined for all  $i, j$  (cf., Section 1.5) by

$$(3.1) \quad G_{ij}(t) = P[N_j(t) > 0 \mid Z_0 = i], \quad P_{ij}(t) = P[Z_t = j \mid Z_0 = i]$$

for  $t \geq 0$ , and denote their respective L.-S. transforms (when they exist) by  $g_{ij}$  and  $\pi_{ij}$ . Let  $b_{ij}$ ,  $\mu_{ij}$  and  $\eta_i$  be the first moments (possibly infinite) of the mass functions  $F_{ij}$ ,  $G_{ij}$  and  $H_i$  respectively. Notice that for  $i = j$ ,  $G_{ii}$  becomes the mass function (possibly of total variation less than one) representing the probabilistic behavior of the first passage time of the process from state  $i$  into state  $i$ , but that the process need not have left state  $i$  during this recurrence time.

The two quantities defined by (3.1) are closely related both to each other and to the elements of the basic  $\mathcal{Q}$ -matrix of the M.R.P. being studied. In the next three lemmas these relationships are explicitly stated. First of all, the relationship between the  $P_{ij}$ 's and the  $Q_{ij}$ 's is given in

LEMMA 3.1. For all  $t \geq 0$ ,  $s > 0$

$$(3.2) \quad P_{ij}(t) = \delta_{ij} - \sum_{k=1}^m [\delta_{ij} - P_{kj}(t)] * Q_{ik}(t),$$



$$(3.3) \quad \pi_{ij}(s) = \delta_{ij} - \sum_{k=1}^m [\delta_{ik} - \pi_{kj}(s)] q_{ik}(s).$$

[Throughout this paper,  $\delta_{ij}$  is used to indicate both the Kronecker delta and the function  $\delta_{ij}U_0(\cdot)$ , where  $U_c$  is the d.f. with unit jump at  $c$ , whose domain (possibly restricted, as it is in (3.3)) is determined by the context. A similar remark applies to any real constant.]

PROOF. Define for all  $t \geq 0$ ,  $n \geq 0$ ,

$$(3.4) \quad P_{ij}(t; n) = P[Z_t = j, N(t) = n \mid Z_0 = i].$$

By Lemma I.4.1.,  $N(t) < \infty$  a.s. since  $m < \infty$ , and so  $P_{ij}(t) = \sum_{n=0}^{\infty} P_{ij}(t; n)$ . The quantities  $P_{ij}(t; n)$  are straightforwardly shown to satisfy  $P_{ij}(t; 0) = \delta_{ij}[1 - H_i(t)]$  and, for  $n > 0$ ,

$$(3.5) \quad P_{ij}(t; n) = \sum_{k=1}^m P_{kj}(t; n-1) * Q_{ik}(t).$$

Upon summing over  $n$  in these last two equations, (3.2) is obtained immediately. (3.3) then follows upon taking L.-S. transforms of the former.

The analogous relationship between the  $G_{ij}$ 's and the  $Q_{ij}$ 's is given without proof in

LEMMA 3.2. For  $t \geq 0$ ,  $s > 0$ ,

$$(3.6) \quad G_{ij}(t) = \sum_{k=1}^m G_{kj}(t) * Q_{ik}(t) + [1 - G_{jj}(t)] * Q_{ij}(t)$$

$$(3.7) \quad g_{ij}(s) = \sum_{k=1}^m g_{kj}(s) q_{ik}(s) + [1 - g_{jj}(s)] q_{ij}(s).$$

Between the  $G_{ij}$ 's and the  $P_{ij}$ 's, there is a particularly tractable relationship, as given in

THEOREM 3.1. For  $t \geq 0$ ,  $s > 0$ ,

$$(3.8) \quad P_{ij}(t) = P_{jj}(t) * G_{ij}(t) + \delta_{ij}[1 - H_i(t)],$$

$$(3.9) \quad \pi_{ij}(s) = \pi_{jj}(s) g_{ij}(s), \quad (i \neq j); \quad \pi_{jj}(s) = \frac{1 - h_j(s)}{1 - g_{jj}(s)}.$$

PROOF. It suffices to remark roughly that for  $i \neq j$ ,  $P_{ij}(t)$  is the probability of reaching state  $j$  for the first time before  $t$  (according to  $G_{ij}$ ) and then, in the remaining time, ending up in state  $j$  (according to  $P_{jj}$ ). In case  $i = j$ , one must add to the above the probability that no transition occurs in  $(0, t]$ , namely,  $1 - H_i(t)$ . (3.9) is an immediate consequence of (3.8).

Define

$$(3.10) \quad D_i(t) = 1 - P[Z_u = i, 0 \leq u < t \mid Z_0 = i].$$

Then  $D_i$  is a mass function representing the duration of time that the Z-process

remains in state  $i$ . It is clear that  $D_i = H_i$  in case  $p_{ii} = 0$ . In particular,  $D_i = H_i$  for an S.-M.P. One may easily prove

LEMMA 3.3. For  $t \geq 0$ ,  $s > 0$ ,

$$(3.11) \quad \begin{aligned} D_i(t) &= [H_i(t) - Q_{ii}(t)][1 - Q_{ii}(t)]^{(-1)} \\ d_i(s) &= [h_i(s) - q_{ii}(s)][1 - q_{ii}(s)]^{-1}. \end{aligned}$$

Notice that  $D_i$  must have total variation equal to either zero or one, and that it equals zero if and only if  $p_{ii} = 1$ . In this case one would say that state  $i$  is an absorbing state.

The explicit evaluation of  $P_{jj}$  in terms of  $H_j$  and  $G_{jj}$  determined by (3.9) may be used to characterize a recurrent state in terms of the integrability of  $P_{jj}$ . The reader is referred to Section I.5 for the definition of a recurrent state in an M.R.P.

THEOREM 3.2. If  $\eta_j < \infty$ , then state  $j$  is recurrent if and only if

$$(3.12) \quad \int_0^\infty P_{jj}(t) dt = \infty.$$

PROOF. By its definition in (3.1) one has

$$(3.13) \quad \lim_{s \rightarrow 0} s^{-1} \pi_{jj}(s) = \lim_{s \rightarrow 0} \int_0^\infty e^{-st} P_{jj}(t) dt = \int_0^\infty P_{jj}(t) dt$$

as a consequence of the Lebesgue Monotone Convergence Theorem, whether or not the limit is finite. From Theorem 3.1, one obtains, when  $\eta_j < \infty$ ,

$$(3.14) \quad \lim_{s \rightarrow 0} s^{-1} \pi_{jj}(s) = \lim_{s \rightarrow 0} \frac{1 - h_j(s)}{s} [1 - q_{jj}(s)]^{-1} = \eta_j [1 - G_{jj}(\infty)]^{-1}.$$

Therefore (3.13) and (3.14) together imply that state  $j$  is recurrent (i.e.,  $G_{jj}(\infty) = 1$ ) if and only if (3.12) holds, as required.

Note that because of Theorem I.5.1(a), it follows from the above theorem, that if  $\eta_j < \infty$ , then

$$\int_0^\infty P_{jj}(t) dt < \infty \quad \text{if and only if} \quad \sum_{n=0}^\infty p_{jj}^{(n)} < \infty$$

where  $\mathbf{P}^n = (p_{ij}^{(n)})$ . It should also be emphasized that in the proof of Theorem 3.2, the following stronger result is obtained, namely, that if  $\eta_j < \infty$

$$G_{jj}(\infty) = 1 - \eta_j \left\{ \int_0^\infty P_{jj}(t) dt \right\}^{-1}.$$

**4. The matrices  $\Phi$  and  $\mathcal{G}$ .** In this section, the relationships given in Lemmas 3.1 and 3.2 will be solved, with explicit expressions for the  $P_{ij}$  and the  $G_{ij}$  in terms of the  $Q_{ij}$  being given in matrix notation. Theoretically, therefore, for any M.R.P., the  $G_{ij}$  and the  $P_{ij}$  may be uniquely determined from the  $Q_{ij}$ .

Consider the matrix-valued functions defined by

$$\begin{aligned}\mathcal{P} &= (P_{ij}), & \pi &= (\pi_{ij}) \\ \mathcal{G} &= (G_{ij}), & g &= (g_{ij}) \\ \mathcal{K} &= (\delta_{ij}H_i), & k &= (\delta_{ij}h_{ij}).\end{aligned}$$

Define a convolution operation on matrix-valued functions (whenever the definition is valid) by  $\mathcal{K} * \mathcal{L} = (\sum_{k=1}^n K_{ik} * L_{kj})$  which is formally the same as regular matrix multiplication except that the usual numerical product is replaced with convolution. Let  $\mathbf{I}$  denote either the identity matrix  $(\delta_{ij})$  or the unit-step matrix-valued function  $(\delta_{ij}U_0(\cdot))$ . The domains of the various functions will be clear from their contexts. For an arbitrary matrix-valued function  $\mathcal{K}$ , set  $\mathcal{K}^{(0)} = \mathbf{I}$  and define  $\mathcal{K}^{(n)} = \mathcal{K}^{(n-1)} * \mathcal{K}$  for  $n > 0$  and  $(\mathbf{I} - \mathcal{K})^{(-1)} = \sum_{n=0}^{\infty} \mathcal{K}^{(n)}$  whenever the series converges. An explicit expression for  $\mathcal{P}(\pi)$  in terms of  $\mathcal{Q}$  and  $\mathcal{K}(\mathfrak{q}$  and  $k$ ) is given in

THEOREM 4.1.

$$(4.1) \quad \mathcal{P} = (\mathbf{I} - \mathcal{Q})^{(-1)} * (\mathbf{I} - \mathcal{K}), \quad \pi = (\mathbf{I} - \mathfrak{q})^{-1}(\mathbf{I} - k),$$

the latter equation being defined over  $(0, \infty)$ , while  $\pi(0) = 0$ .

PROOF. It suffices to show that  $(P_{ij}(\cdot; n) = \mathcal{Q}^{(n)} * (\mathbf{I} - \mathcal{K})$  where  $P_{ij}(\cdot; n)$  is as defined by (3.4), since then in view of the fact that  $N(t) < \infty$  a.s., (4.1) follows immediately upon summation over  $n$ . It is easily checked recursively that  $P_{ij}(\cdot; n) * Q_{kl} = Q_{kl} * P_{ij}(\cdot; n)$ . It therefore follows that  $(P_{ij}(\cdot; 0)) = \mathbf{I} - \mathcal{K}$  and  $(P_{ij}(\cdot; n)) = \mathcal{Q} * (P_{ij}(\cdot; n-1))$ ,  $(n > 0)$ . Consequently, one obtains

$$(4.2) \quad \mathcal{P} = \sum_{n=0}^{\infty} \mathcal{Q}^{(n)} * (\mathbf{I} - \mathcal{K}),$$

which immediately yields

$$(4.3) \quad \pi(s) = \sum_{n=0}^{\infty} [\mathfrak{q}(s)]^n [\mathbf{I} - k(s)],$$

for  $s > 0$ . To verify (4.1), let  $s$  be a fixed positive number and define  $c_s = \max_{i,j} q_{ij}(s)$ . Since by (i) of Definition I.3.1. each mass function  $Q_{ij}$  satisfies  $Q_{ij}(0) = 0$ , and since  $m < \infty$ , one has  $c_s < 1$ . Therefore  $0 \leq \mathfrak{q}^n(s) \leq c_s^n \mathbf{1}$  where the inequality is termwise and where  $\mathbf{1}$  denotes the  $m \times m$  matrix in which each term equals 1. Consequently the series in (4.3) converges. Moreover  $[\mathbf{I} - \mathfrak{q}(s)] \sum_{n=0}^{\infty} \mathfrak{q}^n(s) = \mathbf{I}$ , and hence the inverse  $(\mathbf{I} - \mathfrak{q})^{-1} = \sum \mathfrak{q}^n$  is well defined for all  $s > 0$ . That  $\pi(0) = \mathbf{I}$  follows directly from the definition, (3.1).

It should be remarked here that equation (3.3) of Lemma 3.1 may be rewritten in matrix notation as  $\pi = \mathbf{I} - k + \mathfrak{q}\pi$ . Equation (4.1) of the above theorem may then be considered as an immediate consequence of Lemma 3.1, once the non-singularity of  $\mathbf{I} - \mathfrak{q}$  is demonstrated.

For any  $m \times m$  matrix (or matrix-valued function)  $\mathbf{A} = (a_{ij})$ , define the

diagonal and off-diagonal parts of  $\mathbf{A}$  by

$${}_d\mathbf{A} = (\delta_{ij}a_{ij}), \quad {}_o\mathbf{A} = \mathbf{A} - {}_d\mathbf{A}.$$

With this notation one may rewrite (3.7) of Lemma 3.2 and (3.9) of Theorem 3.1, respectively, as

$$(4.4) \quad \mathfrak{g} = \mathfrak{q}(\mathbf{I} + {}_o\mathfrak{g})$$

and

$$(4.5) \quad \pi = (\mathbf{I} + {}_o\mathfrak{g})(\mathbf{I} - {}_d\mathfrak{g})^{-1}(\mathbf{I} - \mathfrak{h}).$$

Equivalently, because of (3.9), (4.5) may be rewritten as  $\pi({}_d\pi)^{-1} = \mathbf{I} + {}_o\mathfrak{g}$  since for  $s > 0$ ,  $\pi_{jj}(s) > 0$  for all  $j$ . Consequently, substitution of this in (4.4) leads, as a result of Theorem 4.1, to the proof of

THEOREM 4.2. As defined on  $(0, \infty)$

$$(4.6) \quad \mathfrak{g} = \mathfrak{q}\pi({}_d\pi)^{-1} = \mathfrak{q}(\mathbf{I} - \mathfrak{q})^{-1}\{ {}_d[(\mathbf{I} - \mathfrak{q})^{-1}] \}^{-1}.$$

An immediate consequence of Theorem 4.2 is that a formula can be given for the mean recurrence times  $\mu_{ii}$  as defined after (3.1). Set  $\mathbf{u} = (\mu_{ij})$ . Clearly  $\mathbf{u} = \lim_{s \rightarrow 0} s^{-1}(\mathbf{I} - \mathfrak{g})$ , and so from (4.6) it can be shown that

$$(4.7) \quad {}_d\mathbf{u} = \lim_{s \rightarrow 0} \{ {}_d[s(\mathbf{I} - \mathfrak{q})^{-1}] \}^{-1}$$

where one must interpret  $1/0 = \infty$ .

**5. The probability distribution of  $N_j(t)$ .** It is possible to obtain a system of integral equations for the probability distributions of the r.v.'s  $N_j(t)$  ( $1 \leq j \leq m$ ). Explicit solutions of these equations are derived in terms of double generating functions of these probabilities. Theoretically, therefore, the probabilities are determined and, in particular, the moments of  $N_j(t)$  may then be obtained in the usual way.

Define for all  $1 \leq i, j \leq m$ ,  $k \geq 0$ ,  $s, t \geq 0$ , and  $|z| \leq 1$ ,

$$(5.1) \quad v_{ij}(k; t) = P[N_j(t) = k \mid Z_0 = i], \quad \phi_{ij}(z; t) = \sum_{k=0}^{\infty} z^k v_{ij}(k; t),$$

$$\psi_{ij}(z; s) = \int_{0-}^{\infty} e^{-st} d_t \phi_{ij}(z; t),$$

and  $\mathbf{V}_k = (v_{ij}(k; \cdot))$ ,  $\Psi_s = (\psi_{ij}(z; \cdot))$ .

It is immediately seen that the  $v_{ij}$  can be expressed either in terms of the  $Q_{ij}$ 's or in terms of the  $G_{ij}$  by means of the following integro-difference equations. These expressions are self explanatory and require no proofs. For  $t \geq 0$

$$(5.2) \quad \begin{aligned} v_{ij}(k; t) &= \sum_{r \neq j} v_{rj}(k; t) * Q_{ir}(t) + v_{jj}(k-1; t) * Q_{ij}(t), \quad (k > 0), \\ v_{ij}(0; t) &= \sum_{r \neq j} v_{rj}(0; t) * Q_{ir}(t) + 1 - H_i(t) \end{aligned}$$

and

$$(5.3) \quad v_{ij}(k; t) = v_{jj}(k-1, t) * G_{ij}(t), \quad v_{ij}(0; t) = 1 - G_{ij}(t), \quad (k > 0).$$

The solutions to (5.3) are easily seen to be

$$(5.4) \quad \begin{aligned} v_{ij}(k; t) &= G_{ij}(t) * G_{jj}^{(k-1)}(t) * [1 - G_{jj}(t)], & (k > 0), \\ v_{ij}(0; t) &= 1 - G_{ij}(t). \end{aligned}$$

Actually (5.4) may be viewed as a known result in Renewal Theory. By (5.4), the  $v_{ij}$  are explicitly expressed in terms of the  $G_{ij}$ , which in turn have been expressed in Theorem 4.2 in terms of the basic  $Q_{ij}$ 's. This relationship may be more simply expressed by means of generating functions. Thus from (5.4), one obtains for  $|z| \leq 1$ ,

$$\phi_{ij}(z; t) = 1 - G_{ij}(t) + zG_{ij}(t) * [1 - zG_{jj}(t)]^{(-1)} * [1 - G_{jj}(t)],$$

and so for  $s \geq 0$

$$\psi_{ij}(z; s) = 1 - g_{ij}(s) + zg_{ij}(s)[1 - g_{jj}(s)][1 - zg_{jj}(s)]^{-1}.$$

From the last expression, one obtains

THEOREM 5.1. As defined over  $(0, \infty)$  for  $|z| \leq 1$ , one has

$$(5.5) \quad \Psi_z = \mathbf{1} - (1 - z)g(\mathbf{I} - z_d g)^{-1}.$$

Substitution for  $g$  in terms of  $q$  may be made in (5.5) by Theorem 4.2, thus giving an explicit expression for  $\Psi$  in terms of  $q$  (cf., Corollary 5.1 below). Notice that  $\Psi_0 = \mathbf{1} - g$  and  $\Psi_1 = \mathbf{1}$ , as required.

An alternative derivation of (5.5) is to bypass the explicit expression (5.4) and to derive the analogous equations to (5.3) for the generating functions. By so doing, one would obtain directly the matrix equation

$$(5.6) \quad \Psi_z - zg_d \Psi_z = \mathbf{1} - g.$$

Now for any square matrices  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{D}$  for which  $c_{ii} \neq 1$  for every  $i$ , the solution for  $\mathbf{B}$  in the equation  $\mathbf{B} - \mathbf{C}_d \mathbf{B} = \mathbf{D}$  is easily checked to be

$$(5.7) \quad \mathbf{B} = \mathbf{D} + \mathbf{C}(\mathbf{I} - {}_d \mathbf{C})^{-1} {}_d \mathbf{D},$$

since the given equation implies that  $(\mathbf{I} - {}_d \mathbf{C}) {}_d \mathbf{B} = {}_d \mathbf{D}$ , and the assumption permits the taking of the inverse of  $\mathbf{I} - {}_d \mathbf{C}$ . Theorem 5.1 may then be obtained by applying (5.7) to (5.6).

In terms of generating functions the equation (5.2) becomes

$$\psi_{ij}(z; s) = \sum_{r=1}^m q_{ir}(s) \psi_{rj}(z; s) - (1 - z)q_{ij}(s) \psi_{jj}(z; s) + 1 - h_i(s),$$

or in matrix notation

$$(5.8) \quad (\mathbf{I} - q) \Psi_z + (1 - z)q {}_d \Psi_z = (\mathbf{I} - q)\mathbf{1}.$$

As in the proof of Theorem 4.1,  $(\mathbf{I} - \mathbf{q})$  is non-singular. Therefore, (5.8) may be rewritten as

$$\Psi_z + (1 - z)(\mathbf{I} - \mathbf{q})^{-1}\mathbf{q} \, {}_d\Psi_z = \mathbf{1},$$

whose solution, following (5.7), is given in

COROLLARY 5.1. *As defined over  $(0, \infty)$  for  $|z| \leq 1$ , one has*

$$(5.9) \quad \Psi_z = \mathbf{1} - (1 - z)(\mathbf{I} - \mathbf{q})^{-1}\mathbf{q}[z\mathbf{I} + (1 - z){}_d\{(\mathbf{I} - \mathbf{q})^{-1}\}]^{-1}.$$

This is indeed a corollary of Theorem 5.1, since, as mentioned earlier, it can be obtained from (5.5) by a substitution of (4.6). However, the above more direct approach is of interest in its own right.

Although the explicit result given in (5.9) is somewhat complex in appearance, it should be observed that it implies that when computing  $\Psi_z$  from  $\mathbf{q}$ , only one major computation must be made, namely the inversion of  $(\mathbf{I} - \mathbf{q})$ . This remark applies also to the results of the preceding section. The reader should also note that Theorem 4.2 may be viewed as a corollary to (5.9), since by definition  $\mathbf{1} - \Psi_0 = \mathbf{g}$ .

Clearly the moments of  $N_j(t)$ , or more precisely, the L.-S. transforms of these moments, may be obtained by successive differentiations of  $\Psi_z$ . In particular, the L.-S. transform of the expectation of  $N_j(t)$  is readily obtained from (5.9) because of the special form of  $\Psi_z$ . Define for  $t \geq 0$ ,  $s > 0$

$$(5.10) \quad M_{ij}(t) = E[N_j(t) \mid Z_0 = i], \quad m_{ij}(s) = \int_0^\infty e^{-st} dM_{ij}(t),$$

and set  $\mathfrak{M} = (M_{ij}(\cdot))$ ,  $\mathfrak{m} = (m_{ij}(\cdot))$ . We shall call  $\mathfrak{M}$  the *Renewal function* of the process. Clearly

$$\mathfrak{m} = (z - 1)^{-1}(\Psi_z - \mathbf{1})|_{z=1}.$$

From (5.5) and (5.9), one then obtains, respectively,

$$\mathfrak{m} = \mathbf{g}(\mathbf{I} - {}_d\mathbf{g})^{-1} = \mathbf{q}(\mathbf{I} - \mathbf{q})^{-1},$$

thus proving

THEOREM 5.2. *For a M.R.P. with  $m < \infty$ , the (conditional) expectations of  $\mathbf{N}(t)$  satisfy*

$$(5.11) \quad \mathfrak{M} = \mathbf{Q} * (\mathbf{I} - \mathbf{Q})^{(-1)} = (\mathbf{I} - \mathbf{Q})^{(-1)} - \mathbf{I} \quad \text{on } [0, \infty)$$

and

$$(5.12) \quad \mathfrak{m} = \mathbf{q}(\mathbf{I} - \mathbf{q})^{-1} = (\mathbf{I} - \mathbf{q})^{-1} - \mathbf{I} \quad \text{on } (0, \infty).$$

This is, in several respects, a very important result, if not at first sight amazing. First of all, it implies that a knowledge of  $\mathfrak{M}$  is equivalent to a knowledge of  $\mathbf{Q}$ . That is, an M.R.P. is equally as well determined by  $(\mathfrak{m}, \mathbf{A}, \mathfrak{M})$  as by  $(\mathfrak{m}, \mathbf{A}, \mathbf{Q})$ . Because of known results in Renewal theory, the moments of the  $G_{ii}$  may be determined by a knowledge of the asymptotic behavior of the  $M_{ii}$ . In the next section we shall briefly consider some special cases of M.R.P.'s, one of which is



the case of a Renewal process (the case of  $m = 1$ ). The striking similarity between (5.11) and (5.12) and the corresponding known results for Renewal processes will then become apparent. It seems to this author that the suitability of the name Markov Renewal processes for the stochastic processes being studied in these papers, is best supported by this similarity, together with the ease with which Renewal theory yields limiting results for these processes, as is demonstrated in Section 7.

Theorem 3.2 together with (5.12) shows that one may write

$$(5.13) \quad \mathfrak{M} = \mathfrak{G}(\mathfrak{A}\mathfrak{M} + \mathbf{I}), \quad \mathfrak{G} = \mathfrak{M}(\mathfrak{A}\mathfrak{M} + \mathbf{I})^{(-1)}$$

which implies  $M_{ij} = G_{ij} * M_{jj} + G_{ij}$ .

Because of the basic nature of Theorem 5.2, it is desirable to consider the following more direct proof which affords a much clearer insight into this relationship between  $\mathfrak{M}$  and  $\mathfrak{Q}$ , making it intuitive and natural, rather than "amazing". From the definition of an M.R.P. (in particular, I.(3.6)) it follows that

$$(5.14) \quad P[J_n = j, S_n \leq t | Z_0 = i] = \sum_{g_n, i, j}^{n-1} * Q_{\alpha_n \alpha_{n+1}}(t).$$

Hence, either by induction or by recalling the interpretation of elements of  $\mathbf{P}^n$  in Markov Chain theory, one obtains

$$(5.15) \quad (P[J_n = j, S_n \leq t | Z_0 = i]) = \mathfrak{Q}^{(n)} = \{\gamma_{ij}^{(n)}\}.$$

Furthermore, define for each  $j$ , r.v.'s  $\{U_{n,j}; n \geq 1\}$  by  $U_{n,j} = 1$  if  $J_n = j$  and  $S_n \leq t$ , and  $= 0$  otherwise. Clearly  $N_j(t) = \sum_{n=1}^{\infty} U_{n,j}$ . Therefore, since  $(E[U_{n,j} | Z_0 = i]) = \mathfrak{Q}^{(n)}$ , one obtains

$$(5.16) \quad \mathfrak{M} = (E[N_j(t) | Z_0 = i]) = \sum_{n=1}^{\infty} \mathfrak{Q}^{(n)} = (\mathbf{I} - \mathfrak{Q})^{(-1)} - \mathbf{I}$$

as desired.

## 6. Special cases of Markov Renewal processes.

(a) *Markov Chains*. As has been mentioned earlier, an M.R.P. becomes a Markov Chain whenever  $F_{ij} = U_1(\cdot)$  for all  $i, j$ . Hence for a Markov Chain,

$$\begin{aligned} f_{ij}(s) &= e^{-s}, & q_{ij}(s) &= p_{ij}e^{-s}, & h_i(s) &= e^{-s}, \\ f &= e^{-s}\mathbf{1}, & \mathfrak{q} &= e^{-s}\mathbf{P}, & \mathfrak{k} &= e^{-s}\mathbf{I}. \end{aligned}$$

Consider first of all the relationship (3.12) which becomes

$$(6.1) \quad \pi_{jj}(s) = \frac{1 - h_j(s)}{1 - g_{jj}(s)} = \frac{1 - e^{-s}}{1 - g_{jj}(s)}.$$

Set  $z = e^{-s}$ . For purposes of this paragraph alone, introduce the notation

$$F_j(z) \equiv g_{jj}(-\log z) = \sum_{n=1}^{\infty} z^n f_j(n)$$

$$U_j(z) \equiv (1 - z)^{-1} \pi_{jj}(-\log z) = \sum_{n=0}^{\infty} z^n P_{jj}(n)$$

where  $f_j(n) = G_{jj}(n) - G_{jj}(n-1)$  is the probability that state  $j$  is reentered for the first time at time  $n$ . Upon rewriting (6.1) with this notation, one obtains  $U_j(z) = [1 - F_j(z)]^{-1}$ , which is a very well known relationship for Markov Chains (for example, see Feller [6] pp. 285, 352).

Consider now equation (4.1). For a Markov Chain it becomes

$$\Phi = \sum_{k=0}^{\infty} U_k(\cdot) \mathbf{P}^k * [1 - U_1(\cdot)], \quad \pi(s) = (\mathbf{I} - e^{-s}\mathbf{P})^{-1}(1 - e^{-s}).$$

In particular, one obtains the known result that

$$\Phi(n) = \sum_{k=0}^{\infty} [U_k(n) - U_{k+1}(n)] \mathbf{P}^k = \mathbf{P}^n.$$

Moreover, equation (4.6) of Theorem 4.2 becomes

$$g(s) = e^{-s}\mathbf{P}(\mathbf{I} - e^{-s}\mathbf{P})^{-1}\{d[(\mathbf{I} - e^{-s}\mathbf{P})^{-1}]\}^{-1}$$

while (4.7) becomes

$$(6.2) \quad d\mathbf{y} = \lim_{s \rightarrow 0} \{d[s(\mathbf{I} - e^{-s}\mathbf{P})^{-1}]\}^{-1}.$$

By a straightforward generalization of well known Abelian and Tauberian theorems for series, one obtains for matrices that

$$\lim_{z \rightarrow 1} (1 - z)(\mathbf{I} - z\mathbf{P})^{-1} = \lim_{z \rightarrow 1} \frac{\sum_{k=0}^{\infty} z^k \mathbf{P}^k}{\sum_{k=0}^{\infty} z^k} = \mathbf{L}$$

if and only if

$$(6.3) \quad \lim_{n \rightarrow \infty} n^{-1}(\mathbf{I} + \mathbf{P} + \cdots + \mathbf{P}^n) = \mathbf{L}.$$

Since  $d\mathbf{y}$  obviously exists (possibly with some infinite entries) one deduces from (6.2) the well known ergodic result that  $d\mathbf{y} = (d\mathbf{L})^{-1}$ .

From Theorem 5.2 one obtains for a Markov Chain, that for  $n > 0$ ,  $\mathfrak{M}(n) = \sum_{k=1}^n \mathbf{P}^k$ , while, quite obviously,  $\mathfrak{M}$  is constant over every interval of the form  $[n, n+1)$ .

(b) *Continuous parameter Markov processes with finitely many states.* Such a process is a special case of an M.R.P. for which  $p_{ii} = 0$  and  $F_{ij}(t) = 1 - e^{-\lambda_i t}$  for appropriate finite  $\lambda_i > 0$ . Clearly  $F_i = H_i$ . Setting  $\eta = (\delta_{ij}\eta_i) = (\delta_{ij}\lambda_i^{-1})$ , one may write

$$\mathfrak{q} = (\mathbf{I} + s\eta)^{-1}\mathbf{P},$$

and hence from (4.1) and (5.12) one obtains

$$\begin{aligned} \pi(s) &= [\mathbf{I} - (\mathbf{I} + s\eta)^{-1}\mathbf{P}]^{-1}[\mathbf{I} - (\mathbf{I} + s\eta)^{-1}] \\ \mathfrak{M}(s) + \mathbf{I} &= [\mathbf{I} - (\mathbf{I} + s\eta)\mathbf{P}]^{-1}. \end{aligned}$$

The corresponding expression for  $\Phi$  may also be written down, and from it,

one would expect to be able to derive the semigroup property of  $\mathcal{P}$ . This does not seem to be, however, a very simple deduction.

(c) *Renewal processes.* A Renewal process (R.P.) is defined as a sequence of independent and identically distributed r.v.'s, say  $\{X_n: n \geq 1\}$ . [The reader is referred to the survey paper by Smith [7] for details of Renewal theory, as well as for complete references to the proofs of the theorem stated below.] Equivalently, either the sequence of partial sums  $\{S_n: n \geq 1\}$ , or the process  $\{N(t): t \geq 0\}$  defined as before by  $N(t) = \sup \{k: S_k \leq t\}$ , can be termed a Renewal process. With emphasis upon the latter description, it is easily seen that the family of all R.P.'s and the family of all M.R.P.'s having  $m = 1$  are identical. Most of the above results become either vacuously true (e.g., the results of Sections 3 and 4), or obvious (e.g., the results of Section 5) for R.P.'s. Note that for a R.P.,  $Q_{11} = F_{11}$ . The analogues of Theorems 5.1, 5.2, for example, become, dropping all redundant subscripts,

$$(6.4) \quad v(k; \cdot) = F^{(k)} - F^{(k+1)}, \quad \psi(z; \cdot) = (1-f)(1-zf)^{-1}, \\ M = (1-F)^{(-1)} - 1, \quad \text{and} \quad m = f(1-f)^{-1},$$

all of which can be derived directly with very little effort. It is the similarity between (5.12) and (6.4) that is referred to at the end of Section 5.

As may be seen in [7], Renewal theory goes very much deeper than these finite results, the basic emphasis being on the study of the limiting behavior of  $M(t)$ . In the following section, the limiting stationarity of an M.R.P. will be discussed as an application of the main limit theorem of Renewal theory, due to Blackwell and Smith, which states that if  $k$  is any non-negative, non-increasing, Lebesgue-integrable function defined on  $[0, \infty)$ , then

$$(6.5) \quad k(t) * M(t) \rightarrow \begin{cases} \mu^{-1} \int_0^\infty k(x) dx & \text{if } F \text{ is non-lattice} \\ h\mu^{-1} \sum_{n=0}^\infty k(nh) & \text{if } F \text{ is lattice with span } h, \end{cases}$$

where for the lattice case,  $t \rightarrow \infty$  over multiples of  $h$ .

**7. Stationary probabilities.** In this section we derive the limiting form of a certain d.f. pertaining to an M.R.P., and use this to select the appropriate initial distribution for making the corresponding S.-M.P. stationary. The method of derivation is to compute the pertinent d.f., a useful formula in itself, and to apply the Blackwell-Smith theorem (6.5) to it to obtain its limiting form.<sup>5</sup> In [3], Smith derived the asymptotic form of the probabilities  $P_{ij}(t)$  of an M.R.P. For discrete or continuous parameter Markov processes it is known that such is sufficient for ascertaining the initial distribution which makes the process

<sup>5</sup> As was pointed out by the referee, it is also possible to apply the more general Renewal theorem of Smith (Corollary 2.1 of [3]) upon making a suitable redefinition of the state space.

stationary. However, the problem of obtaining the stationary probabilities of an M.R.P. is not solved by deriving the limits of the  $P_{ij}(t)$ . One must instead consider the problem of finding the limit, as  $t \rightarrow \infty$ , of the probability of being in state  $j$  at time  $t$ , of making the next transition sometime before  $t + x$  and of this next transition being into state  $k$ .

First of all, define formally the probability just referred to, as

$$R_{jk}^{(i)}(x; t) = P[Z_t = j, J_{N(t)+1} = k, S_{N(t)+1} \leq t + x \mid Z_0 = i].$$

Now by (5.14) and (5.15) one may straightforwardly show that

$$\begin{aligned} R_{jk}^{(i)}(x; t) &= \sum_{n=0}^{\infty} P[J_n = j, J_{n+1} = k, S_n \leq t < S_{n+1} \leq t + x \mid Z_0 = i] \\ &= \sum_{n=0}^{\infty} [Q_{jk}(t + x) - Q_{jk}(t)] * Q_{ij}^{(n)}(t). \end{aligned}$$

Hence by Theorem 5.2 and (5.13) one may write

$$\begin{aligned} R_{jk}^{(i)}(x; t) &= [Q_{jk}(t + x) - Q_{jk}(t)] * [M_{ij}(t) + \delta_{ij}U_0(t)] \\ (7.1) \quad &= [Q_{jk}(t + x) - Q_{jk}(t)] * G_{ij}(t) * [M_{jj}(t) + U_0(t)] \\ &\quad + \delta_{ij}[Q_{jk}(t + x) - Q_{jk}(t)]. \end{aligned}$$

Upon applying the Renewal theorem, (6.5), to (7.1) with  $M_{jj} = M$  and  $k(t)$  equal first to  $p_{jk} - Q_{jk}(t)$  and then to  $p_{jk} - Q_{jk}(t + x)$ ,<sup>6</sup> one obtains that if  $G_{jj}$  is a non-lattice d.f. (which may be seen to be equivalent to assuming that  $j$  is recurrent and that not every non-zero  $Q_{ir}$  for  $i, r \in C_j$  is a lattice mass function) and if  $b_{jk} < \infty$

$$\begin{aligned} \lim_{t \rightarrow \infty} R_{jk}^{(i)}(x; t) &= G_{ij}(\infty) \mu_{jj}^{-1} \int_0^{\infty} [Q_{jk}(t + x) - Q_{jk}(t)] dt \\ (7.2) \quad &= G_{ij}(\infty) p_{jk} \mu_{jj}^{-1} \int_0^x [1 - F_{jk}(y)] dy \end{aligned}$$

If  $G_{jj}$  is a lattice d.f. of span  $h$ , (that is,  $j$  is recurrent and all non-zero  $Q_{ir}$  for  $i, r \in C_j$  are lattice mass functions) and if  $b_{jk} < \infty$ , then

$$(7.3) \quad \lim_{n \rightarrow \infty} R_{jk}^{(i)}(x; nh) = G_{ij}(\infty) h \mu_{jj}^{-1} \sum_{n=0}^{[x/h]+1} [1 - F_{jk}(nh)]$$

where  $[y]$  is the largest integer less than  $y$ .

Suppose  $G_{jj}$  has variation less than one, and so is not a d.f. Then, as seen, for example, from (5.13),  $\lim_{t \rightarrow \infty} M_{jj}(t) = [1 - G_{jj}(\infty)]^{-1} < \infty$  and hence it follows directly from (7.1) that  $\lim_{t \rightarrow \infty} R_{jk}^{(i)}(x; t) = 0$  for all  $i, k$  and all  $x \geq 0$ . The above results are summarized in

**THEOREM 7.1.**

(i) If state  $j$  is recurrent and  $b_{jk} < \infty$ , then

<sup>6</sup> It may easily be demonstrated that the function  $G_{ij}$  causes no difficulty in applying (6.5), for which the family of  $k$  functions has been unnecessarily restricted.

$$(7.4) \quad \lim_{t \rightarrow \infty} R_{jk}^{(i)}(x; t) = G_{ij}(\infty) p_{jk} \mu_{jj}^{-1} \int_0^x [1 - F_{jk}(y)] dy,$$

where it is understood that if  $G_{jj}$  is a lattice d.f. then both  $t$  and  $x$  may take on as values only multiples of its span.

(ii) If state  $j$  is transient, then  $\lim_{t \rightarrow \infty} R_{jk}^{(i)}(x; t) = 0$  for all  $i, k$  and all  $x \geq 0$ .

Since  $P_{ij}(t) = \sum_{k=1}^m R_{jk}^{(i)}(\infty; t)$  and  $m < \infty$ , one obtains as a consequence of this theorem, the following result of Smith ([3], Theorem 5).

COROLLARY 7.1. For an M.R.P. for which  $\eta_j < \infty$ ,

$$(7.5) \quad \lim_{t \rightarrow \infty} P_{ij}(t) = G_{ij}(\infty) \eta_j / \mu_{jj},$$

with the understanding that  $t$  takes on only multiples of the span if  $G_{ij}$  is a lattice d.f.

Actually, in [3] the right hand side of (7.5) is given as

$$G_{ij}(\infty) \int_0^\infty x dD_j(x) / \int_0^\infty x dK_{ij}(x)$$

where  $D_j$  is as defined in (3.10) and where

$$K_{ij}(t) = P[Z_u \neq i, Z_v = j \text{ for some } u < v \leq t | Z_0 = i].$$

That is,  $K_{ij}$  is the first passage time d.f. of state  $i$  in the corresponding S.M.P. determined by  $(m, \mathbf{A}, \mathbf{Q}^*)$ , (cf., Section 3 of [1]), in which a transition into state  $i$  from itself is not observed. The equivalence of the two limits follows from (3.11) and the relationships for  $i \neq j$ ;

$$(7.6) \quad P_{ij} = K_{ij} * (1 - K_{jj})^{(-1)} * (1 - D_j), \quad \pi_{ij} = k_{ij}(1 - d_j)(1 - k_{jj})^{-1},$$

$$(7.7) \quad P_{jj} = (1 - K_{jj})^{(-1)} * (1 - D_j), \quad \pi_{jj} = (1 - d_j)(1 - k_j)^{-1},$$

$$(7.8) \quad K_{jj} = (1 - Q_{jj})^{(-1)} * (G_{jj} - Q_{jj}), \quad k_{jj} = (g_{jj} - q_{jj})(1 - q_{jj})^{-1},$$

which are obtained using the methods and results of Section 3.

For a Renewal process ( $m = 1$ ). Theorem 7.1. reduces to a result due essentially to Doob [8] (cf., also Smith [9]). Actually if there were to exist only one recurrent state (and hence one for which  $p_{jj} = 1$ ) then Theorem 7.1 is essentially this result.

Assume throughout the remainder of this section that the M.R.P. under consideration has only one recurrent class,  $C$  say, and that it is a positive class. By Theorem I.5.1.(c) this means that  $\eta_j < \infty$  for all  $j \in C$ . Assume also, for simplicity of notation, that each  $G_{jj}$  is non-lattice for every recurrent state  $j$ . Under these assumptions it follows that for every state  $i$ ,  $G_{ij}(\infty) = 1$  and  $\mu_{jj} < \infty$  whenever  $j \in C$ . Consequently,

$$(7.9) \quad \lim_{t \rightarrow \infty} R_{jk}^{(i)}(x; t) = \begin{cases} p_{jk} \mu_{jj}^{-1} \int_0^x [1 - F_{jk}(y)] dy & \text{if } j \in C \\ 0 & \text{if } j \notin C \end{cases}$$

which limits are independent of  $i$ .

Define now a slightly more general process than an M.R.P., namely one in which the first transition time and state,  $(X_1, J_1)$ , has a d.f. determined by an auxiliary matrix  $\tilde{Q}$  of transition distributions. That is, it may be considered as an M.R.P. with a random origin determined by  $\tilde{Q}$ . In keeping with existing terminology in Renewal theory, define a *general Markov Renewal process* (G.M.R.P.) determined by  $(m, \mathbf{A}, \tilde{Q}, Q)$  as a functional of the process  $\{(J_n, \tilde{X}_n): n \geq 0\}$  in exactly the same way as is an M.R.P. (cf., Section I.3.), with the following different probabilistic structure on the  $(J, \tilde{X})$ -process. Let

$$\begin{aligned} \tilde{X}_0 &= 0 \text{ a.s.}, & P[J_0 = k] &= a_k \\ (7.10) \quad P[J_1 = k, \tilde{X}_1 \leq x \mid J_0] &\stackrel{\text{a.s.}}{=} \tilde{Q}_{J_0, k}(x) \\ P[J_n = k, \tilde{X}_n \leq x \mid J_0, J_1, \tilde{X}_1, \dots, J_{n-1}, \tilde{X}_{n-1}] &\stackrel{\text{a.s.}}{=} Q_{J_{n-1}, k}(x) \end{aligned}$$

for  $n > 1$ . Compare this description particularly with that of Definition I.3.3. A similar definition may be made for a G.S.-M.P.

Define  $\tilde{\mathbf{A}} = (\tilde{a}_1, \dots, \tilde{a}_m)$  with  $\tilde{a}_j = \eta_{jj}^{-1}$ , and  $\tilde{Q} = (\tilde{Q}_{ij})$  with

$$(7.11) \quad \tilde{Q}_{ij}(t) = p_{ij} \eta_i^{-1} \int_0^t [1 - F_{ij}(y)] dy.$$

We wish to show that the G.S.-M.P. determined by  $(m, \tilde{\mathbf{A}}, \tilde{Q}, Q)$ , which shall be denoted as the  $\tilde{Z}$ -process, is a stationary process.

From the definition of the functions  $R_{jk}^{(i)}(x; t)$  it may easily be seen that they satisfy the following recursion relationship for all  $t, s \geq 0$ .

$$\begin{aligned} (7.12) \quad R_{jk}^{(i)}(x; t+s) &= \sum_{r,u=1}^m \int_0^s R_{jk}^{(u)}(x; s-y) d_r R_{ru}^{(i)}(y; t) \\ &\quad + R_{jk}^{(i)}(s+x; t) - R_{jk}^{(i)}(s; t). \end{aligned}$$

Upon defining

$$(7.13) \quad R_{jk}(x) = \lim_{t \rightarrow \infty} R_{jk}^{(i)}(x; t) = \tilde{a}_i \tilde{Q}_{jk}(x)$$

by (7.9) and (7.11), and letting  $t \rightarrow \infty$  in (7.12) one obtains

$$(7.14) \quad R_{jk}(x) = \sum_{r,u=1}^m \int_0^s R_{jk}^{(u)}(x; s-y) dR_{ru}(y) + R_{jk}(s+x) - R_{jk}(s)$$

for all  $x, s \geq 0$  and  $1 \leq j, k \leq m$ . For all  $x, t \geq 0, 1 \leq i, j, k \leq m$ , define

$$\tilde{R}_{jk}^{(i)}(x; t) = P[\tilde{Z}_t = j, J_{\tilde{N}(t)+1} = k, \tilde{S}_{\tilde{N}(t)+1} \leq t+x \mid \tilde{Z}_0 = i],$$

which is to say that  $\tilde{R}_{jk}^{(i)}(x; t)$  is the counterpart of  $R_{jk}^{(i)}(x; t)$  as defined for a G.S.-M.P. For a G.S.-M.P. it may be seen that

$$(7.15) \quad \tilde{R}_{jk}^{(i)}(x; t) = \sum_{u=1}^m \int_0^t R_{jk}^{(u)}(x; t-y) d\tilde{Q}_{iu}(y) + \delta_{ij} [\tilde{Q}_{ik}(t+x) - \tilde{Q}_{ik}(t)].$$

In words this equation states that in order to be in state  $j$  at time  $t$  and make the next transition into state  $k$  before time  $t+x$ , when the process starts in state



$i$ , one must either make a first transition at some time  $y \leq t$  into some arbitrary state  $u$  in accordance with  $\tilde{Q}_{iu}$  and then in the remaining time  $t - y$  end up in state  $j$  and make the next transition into state  $k$  before time  $t + x$ , or, in case  $i = j$ , make the first transition during the time interval  $(t, t + x]$ , and make it into state  $k$ . The integrands in (7.15) are without  $\sim$ 's because after the initial transition has occurred the process behaves like an ordinary S.-M.P. From (7.15) it follows that

$$(7.16) \quad \sum_{i=1}^m a_i R_{jk}^{(i)}(x; t) = \sum_{i, u=1}^m \int_0^t R_{jk}^{(u)}(x; t - y) da_i \tilde{Q}_{iu}(y) + a_j \tilde{Q}_{jk}(t + x) - a_j \tilde{Q}_{jk}(t).$$

For the particular G.S.-M.P. determined by  $(m, \tilde{\mathbf{A}}, \tilde{\mathbf{Q}}, \mathbf{Q})$  with  $\tilde{\mathbf{A}}, \tilde{\mathbf{Q}}$  as defined in the paragraph containing (7.11), (7.16) states, in view of (7.13), that

$$\sum_{i=1}^m \tilde{a}_i \tilde{R}_{jk}^{(i)}(x; t) = \sum_{i, u=1}^m \int_0^t R_{jk}^{(u)}(x; t - y) dR_{iu}(y) + R_{jk}(t + x) - R_{jk}(t).$$

As a consequence of (7.14), as well as of the definition of  $\tilde{R}_{jk}^{(i)}(x; t)$ , it therefore follows that

$$\begin{aligned} P[Z_t = j, J_{\tilde{N}(t)+1} = k, \tilde{S}_{\tilde{N}(t)+1} \leq t + x] &= R_{jk}(x) \\ &= p_{jk} \mu_{jj}^{-1} \int_0^x [1 - F_{jk}(y)] dy \end{aligned}$$

which is independent of  $t$ . In particular  $P[Z_t = j] = \eta_j \mu_{jj}^{-1} = \tilde{a}_j$ .

For each  $t$ , define the three-dimensional r.v.  $W_t = (J_{\tilde{N}(t)}, J_{\tilde{N}(t)+1}, \tilde{S}_{\tilde{N}(t)+1} - t)$  whose coordinates respectively record for a G.S.-M.P. the state it is in at time  $t$ , the state into which the next transition will be made, and the remaining time until the next transition will occur. It should be clear that the  $W$ -process is essentially equivalent to the  $\tilde{Z}$ -process in that almost all sample functions of the one can be determined from the other and vice versa. We now summarize the results of the preceding paragraphs in

**THEOREM 7.2.** Consider a given S.-M.P. determined by  $(m, \mathbf{A}, \mathbf{Q})$  for which there is only one positive class. Define  $\tilde{\mathbf{A}} = (\tilde{a}_1, \dots, \tilde{a}_m)$  with  $\tilde{a}_i = \eta_i \mu_{ii}^{-1}$  and  $\tilde{\mathbf{Q}} = (\tilde{Q}_{ij})$  with

$$\tilde{Q}_{ij}(t) = p_{ij} \eta_i^{-1} \int_0^t [1 - F_{ij}(y)] dy,$$

the limiting transition distribution given in Theorem 7.1. Then the  $W$ -process which corresponds to the G.S.-M.P. determined by  $(m, \tilde{\mathbf{A}}, \tilde{\mathbf{Q}}, \mathbf{Q})$  is a stationary process whose marginal d.f. is given by

$$P[W_t \leq (j, k, x)] = \sum_{\substack{j \leq i \\ k \leq r}} p_{jk} \mu_{jj}^{-1} \int_0^x [1 - F_{jk}(y)] dy.$$

It should be clear that a result corresponding to that given in Theorem 7.2 is possible for a G.M.R.P. Indeed, if one represents the M.R.P. by the associated S.-M.P. as defined in Section I.3, then the result for the G.M.R.P. could be viewed as a special case of Theorem 7.2, generalized to cover the case of infinite  $m$ . Since this paper has dealt only with the case of  $m < \infty$ , we shall leave the analogue of Theorem 7.2 for M.R.P.'s until later.

**8. Examples.** In conclusion, we list briefly some specific examples of M.R.P.'s to indicate the broad scope of applications for this family of stochastic processes. First of all, the special case in which  $Q_{ij} = p_i F_j$  for each  $i$  and  $j$ , the  $p_i$ 's being real numbers and the  $F_j$ 's being d.f.'s, arises in electronic counter theory and is studied in detail in [10]. This special case, by a slight reinterpretation of the sample functions, is seen to be essentially equivalent to the "zero order" M.R.P. in which  $Q_{ij} = Q_i$  for each  $i$  and  $j$ .

A second very important special case of an M.R.P., although in one sense somewhat degenerate, is a zero-one process. That is,  $m = 2$  and  $p_{11} = p_{22} = 0$ . For example, in a queueing model, the server is either in a busy state or in an idle state, and these states are entered alternately. In a counter problem, the counter is either dead or free, and it alternates between these two states. The only "parameters" of such zero-one processes are the two d.f.'s of the duration times of the two states. Of course, one still assumes independence between the successive time periods. For such a process it is clear that  $Q_{12} = F_{12} = G_{12} = H_1 = D_1$ ,  $Q_{21} = F_{21} = G_{21} = H_2 = D_2$ , and  $G_{11} = F_{12} * F_{21} = G_{22}$ . Set  $F_{12} = F_1$ , and  $F_{21} = F_2$ . It therefore follows straightforwardly from Theorem 4.1 that, for  $i = 1, 2$ ,

$$P_{ii}(t) = [1 - F_i(t)][1 - F_1(t) * F_2(t)]^{(-1)}$$

$$P_{2-i,i}(t) = [1 - F_i(t)] * F_{2-i}(t) * [1 - F_1(t) * F_2(t)]^{(-1)}$$

which checks with Theorem 1 of [11]. Zero-one processes arise in many problems, and hence have been studied by various authors with various emphases. Let it suffice here to mention the several papers by Takács (cf. [12] and references contained therein) in which the total time spent in one of the states during a given interval of time is particularly studied. The limiting normality of these "sojourn" times, under general conditions, derived by Takács, will be a corollary of a general Central limit theorem given in [13].

In an M.R.P., the probability distribution of the next state depends only on the present state of the process. An important generalization of an M.R.P. arises if one allows this distribution to depend also upon the time it took for the last transition. A special class of such processes is seen to be the class of 2-dependent stationary processes with non-negative r.v.'s, which fact indicates the different approach which would have to be used in studying the more general processes. The following example of an M.R.P. arises as a "first approximation" to this generalization. Let  $F_1$  and  $F_2$  be d.f.'s. Define, for  $c > 0$ ,  $G_c(x, y) = F_1(x)$  if  $y < c$  and  $= F_2(x)$  if  $y \geq c$ . For the process  $\{X_n; n \geq 1\}$ , set

$$P[X_n \leq x \mid X_{n-1}, \dots, X_1] = G_c(x, X_{n-1}).$$

This process is clearly equivalent to an M.R.P. with two states, where the state indicates the subscript of the d.f.,  $F_1$  or  $F_2$ , which was used last. Therefore, for  $i = 1, 2$ ,

$$Q_{ii}(x) = F_i(\min[x, c]), \quad Q_{ij}(x) = F_i(\max[x, c]) - F_i(c).$$

Such a model and its analogue for arbitrary  $m$  may be applicable in life-testing or behavioural problems. One particular application to an inventory problem is the following. Suppose that state 2 indicates that a new efficient water pumping station is in use while state 1 indicates that an old inefficient auxiliary station is also in use. Suppose that the transition times represent the successive times that it takes for the capacity of a reservoir to fall below a fixed level. We assume here that the reservoir is instantly refilled at these times. For such a model, the above example of an M.R.P. could be used in studying the cost of the pumping system as well as the proportion of times the auxiliary pumping station is used.

#### REFERENCES

- [1] RONALD PYKE, "Markov Renewal processes: definitions and preliminary properties," *Ann. Math. Stat.*, Vol. 32 (1961), pp. 1231-1242.
- [2] PAUL LÉVY, "Systèmes semi-Markoviens à au plus une infinité dénombrable d'états possibles," *Proc. Internat. Congr. Math.*, Amsterdam, Vol. 2 (1954), p. 294; "Processus semi-markoviens," *ibid.*, Vol. 3 (1954), pp. 416-426.
- [3] W. L. SMITH, "Regenerative stochastic processes," *Proc. Roy. Soc., Ser. A*, Vol. 232 (1955), pp. 6-31 [cf., the abstract of this paper in *Proc. Internat. Congr. Math.*, Amsterdam, Vol. 2 (1954), pp. 304-305.]
- [4] LAJOS TAKÁCS, "Some investigations concerning recurrent stochastic processes of a certain type," *Magyar Tud. Akad. Mat. Kutató Int. Kozl.*, Vol. 3 (1954), pp. 115-128 (Hungarian).
- [5] GEORGE H. WEISS, "On a semi-markovian process with a particular application to reliability theory." Navord Report No. 4351 (1956) U. S. Naval Ordnance Lab. White Oak, Maryland.
- [6] WILLIAM FELLER, *An Introduction to Probability Theory*, 2nd Ed., John Wiley and Sons, New York, 1957.
- [7] W. L. SMITH, "Renewal theory and its ramifications," *J. Roy. Stat. Soc., Ser. B*, Vol. 20 (1958), pp. 243-302.
- [8] J. L. DOOB, "Renewal theory from the point of view of the theory of probability," *Trans. Amer. Math. Soc.*, Vol. 63 (1948), pp. 422-438.
- [9] W. L. SMITH, "Asymptotic renewal theorems," *Proc. Roy. Soc. Edinburgh., Ser. A*, Vol. 64 (1954), pp. 9-48.
- [10] RONALD PYKE, "Markov Renewal processes of zero order and their applications to counter theory," Technical Report No. 64 (1960), Contract Nonr-255(52), NR - 342-022, Stanford University. To appear in a volume edited by Samuel Karlin, Stanford University Press, Calif.
- [11] RONALD PYKE, "On Renewal processes related to Type I and Type II counter models," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 737-754.
- [12] L. TAKÁCS, "On a sojourn time problem," *Teor. Veroyatnost. i Primenen Akad. Nauk*, Vol. 3 (1958), pp. 61-69.
- [13] RONALD PYKE, "Limit theorems for Markov Renewal processes," Technical Report No. 24 (1961) Contract Nonr-266(59), (042-205), Columbia University.

# A CONVEXITY PROPERTY IN THE THEORY OF RANDOM VARIABLES DEFINED ON A FINITE MARKOV CHAIN

H. D. MILLER<sup>1</sup>

*Statistical Laboratory, University of Cambridge*

**1. Summary.** Let  $P = (p_{jk})$  be the transition matrix of an ergodic, finite Markov chain with no cyclically moving sub-classes. For each possible transition  $(j, k)$ , let  $H_{jk}(x)$  be a distribution function admitting a moment generating function  $f_{jk}(t)$  in an interval surrounding  $t = 0$ . The matrix  $P(t) = \{p_{jk}f_{jk}(t)\}$  is of interest in the study of the random variable  $S_n = X_1 + \dots + X_n$ , where  $X_m$  has the distribution  $H_{jk}(x)$  if the  $m$ th transition takes the chain from state  $j$  to state  $k$ . The matrix  $P(t)$  is non-negative and therefore possesses a maximal positive eigenvalue  $\alpha_1(t)$ , which is shown to be a convex function of  $t$ . As an application of the convexity property, we obtain an asymptotic expression for the probability of tail values of the sum  $S_n$ , in the case where the  $X_m$  are integral random variables.

The results are related to those of Blackwell and Hodges [1], whose methods are followed closely in Section 5, and Volkov [4], [5], who treats in detail the case of integer-valued functions of the state of the chain, i.e., the case  $f_{jk}(t) = \exp(\beta_k t)$  ( $\beta_k$  integral).

**2. Introduction and notation.** Let  $k_m (m = 0, 1, 2, \dots)$  be the state at time  $m$  of a finite  $N$ -state ergodic Markov chain with no cyclically moving subclasses and with transition matrix  $P = (p_{jk})$ , where  $p_{jk} = \Pr(k_m = k | k_{m-1} = j)$ ,  $j, k = 1, \dots, N$ . The distribution of  $k_0$  is unspecified, since we shall mostly deal with probabilities conditional on  $k_0$ . It follows that  $P$  is a non-negative, primitive and irreducible matrix. Let  $H_{jk}(x)$  be a distribution function associated with the transition  $(j, k)$  ( $p_{jk} \neq 0$ ) and let  $f_{jk}(t)$  be the corresponding moment generating function, i.e.,

$$f_{jk}(t) = \int_{-\infty}^{\infty} e^{tx} dH_{jk}(x).$$

We shall suppose that each  $f_{jk}(t)$  is analytic in a strip which strictly contains the imaginary axis of the complex  $t$ -plane. There will therefore be a maximal strip

$$(2.1) \quad u_0 < \operatorname{Re}(t) < u'_0 \quad (-\infty \leq u_0 < 0 < u'_0 \leq \infty),$$

in which all the  $f_{jk}(t)$  are analytic.

Let  $X_m, m = 1, 2, \dots$ , be a random variable having the distribution  $H_{jk}(x)$  if  $k_{m-1} = j$  and  $k_m = k$ , i.e., if the  $m$ th transition is  $(j, k)$ , and let

Received October 18, 1960; revised February 20, 1961.

<sup>1</sup> Present address: Birkbeck College, London

$S_n = X_1 + \cdots + X_n$ . Let  $P(t)$  be the matrix  $\{p_{jk}f_{jk}(t)\}$  and let

$$(2.2) \quad \{P(t)\}^n = \{p_{jk}^{(n)}f_{jk}^{(n)}(t)\},$$

where  $P^n = \{p_{jk}^{(n)}\}$ . Then  $f_{jk}^{(n)}(t)$  is the moment generating function of  $S_n$  conditional on the  $n$ -stage transition from state  $j$  at time 0 to state  $k$  at time  $n$ . Thus

$$(2.3) \quad f_{jk}^{(n)}(t) = E\{\exp(tS_n) \mid k_0 = j, k_n = k\}.$$

For real  $t$ , the matrix  $P(t)$  is non-negative and therefore it has a maximal positive eigenvalue, the Perron root, which we denote by  $\alpha_1(t)$ . Thus  $\alpha_1(0) = 1$ , and, for real  $t$ ,  $\alpha_1(t)$  has the properties (i)  $\alpha_1(t) > 0$ , (ii)  $\alpha_1(t) > |\alpha_j(t)|$ , where  $\alpha_j(t)$ ,  $j = 2, 3, \dots, N$ , are the remaining eigenvalues of  $P(t)$ .

We shall say that  $f(t)$  is a degenerate moment generating function if it is of the form  $e^{\beta t}$  ( $\beta$  real) and we shall say that  $P(t)$  is degenerate if it is of the form

$$(2.4) \quad P(t) = e^{\beta t} D(t) P \{D(t)\}^{-1},$$

where  $D(t)$  is a diagonal matrix of degenerate moment generating functions. If  $P(t)$  is degenerate, then the sum  $S_n$  is also degenerate in the sense that given  $k_0 = j$ ,  $k_n = k$ ,  $S_n$  is deterministic and of the form  $S_n = n\beta + \beta_j - \beta_k$ , where  $D(t) = \text{diag}\{\exp(\beta_j t)\}$ .

Let  $(p_k)$ ,  $k = 1, \dots, N$ , be the unique ergodic distribution associated with  $P$ . Then, if  $k_0$  has the distribution  $(p_k)$  and if we take expectation unconditional on  $k_1$ , it is easy to show that  $E(X_1) = \alpha_1'(0)$ . Thus  $\alpha_1'(0)$  is a measure of the ultimate drift of  $S_n$ .

**3. Some properties of non-negative square matrices.** For the sake of clarity we quote the following properties of non-negative square matrices from the paper of Debreu and Herstein [3].

(a) Let  $A \geq 0$  be an irreducible (indecomposable) square matrix, and let  $\alpha_1$  be its maximal positive eigenvalue. Then  $\alpha_1$  is a simple root of the equation  $|\alpha I - A| = 0$ , and there exist strictly positive left and right eigenvectors corresponding to  $\alpha_1$ . If  $\sigma$  is any other eigenvalue of  $A$ , then  $|\sigma| \leq \alpha_1$ , and if  $|\sigma| < \alpha_1$  then  $A$  is said to be primitive.

(b) A finite stochastic matrix is the transition matrix of a Markov chain which is ergodic and without cyclically moving sub-classes if and only if it is primitive and irreducible.

(c) Let  $B = (b_{jk})$  be a square matrix with complex elements, and let  $B^* = (\overline{b_{jk}})$ . If  $\beta$  is any eigenvalue of  $B$ , if  $A (\geq 0)$  is irreducible, and if  $B^* \leq A$  then  $|\beta| \leq \alpha_1$ . Moreover  $|\beta| = \alpha_1$  and  $B^* \leq A$  together imply that  $B^* = A$ ; if  $\beta = \alpha_1 e^{i\phi}$ , then  $B = e^{i\phi} D^{-1} A D$  where  $D^* = I$ .

In addition we state the following lemma which we need in Sections 4 and 5. It is an immediate consequence of the left and right eigenvector relations.

**LEMMA 3.1.** Let  $A$  be a non-negative, primitive and irreducible matrix of order  $N \times N$ . Let  $\alpha_1$  be its maximal positive eigenvalue with corresponding left and right positive eigenvectors  $y = (y_j)$  and  $x = (x_j)$  respectively, such that  $yx = 1$ . Let

$X = \text{diag}(x_j)$ . Then the matrix  $\alpha_1^{-1}X^{-1}AX$  is a primitive, irreducible, stochastic matrix with limiting probability vector  $(x_j/y_j)$ .

**4. The properties of  $\alpha_1(t)$ .** Let  $t = u + iv$  ( $u, v$  real). Then for  $t$  lying in the strip (2.1), the  $f_{jk}(t)$  and  $P(t)$  satisfy the following conditions:

$$(4.1) \quad \begin{array}{lll} \text{(i)} f_{jk}(u) > 0, & \text{(ii)} |f_{jk}(t)| \leq f_{jk}(u), & \text{(iii)} f_{jk}(0) = 1, \\ & \text{(iv)} P(u) \geq 0, & \text{(v)} \{P(t)\}^* \leq P(u), \end{array}$$

where, in (v), we use the notation of Section 3(c).

**THEOREM 1.**

(a) The function  $\alpha_1(t)$  is regular at each point  $t = u$  of the real axis in the strip (2.1).

(b) An eigenvalue of  $P(t)$  is of the form  $e^{\beta t}$  ( $\beta$  real) if and only if  $P(t)$  is degenerate, i.e., of the form (2.4).

(c) In the strip (2.1) we have

$$\alpha_1(u) \geq |\alpha_j(t)| \quad (j = 2, 3, \dots, N; t = u + iv)$$

**PROOF.**

(a) Since for each real  $t$ ,  $\alpha_1(t)$  is a simple root of the determinantal equation  $|\alpha I - P(t)| = 0$ , and since  $|\alpha I - P(t)|$  is an analytic function of the two complex variables  $\alpha$  and  $t$ , the result follows from the implicit function theorem for analytic functions (Bochner and Martin [2], p. 39).

(b) If  $P(t)$  is of the form (2.4) then clearly  $\alpha_1(t) = e^{\beta t}$ . If  $e^{\beta t}$  is an eigenvalue of  $P(t)$ , then we put  $t = iv$  ( $v$  real), and it follows from (4.1) (v) and Section 3(c) that  $P(iv) = e^{i\beta v} D(v) P\{D(v)\}^{-1}$ , where  $\{D(v)\}^* = I$ . Thus  $|f_{jk}^{(n)}(iv)| = 1$  for each  $j, k$  and  $n$  for which  $p_{jk}^{(n)} > 0$ , and since  $f_{jk}^{(n)}(iv)$  is a characteristic function, we must have  $D(v) = \text{diag}\{\exp(i\beta_j v)\}$  ( $\beta_j$  real  $j = 1, \dots, N$ ). Hence  $P(t)$  is degenerate.

(c) The inequalities follow from (4.1) (v) and Section 3(c).

**THEOREM 2.** If  $P(t)$  is not degenerate, then  $\alpha_1(t)$  ( $t$  real) is a strictly convex function of  $t$ .

**PROOF.** We have the factorization

$$(4.2) \quad |\alpha I - P(t)| = \alpha^N \{1 - \alpha^{-1} \alpha_1(t)\} \{1 - \alpha^{-1} \alpha_2(t)\} \cdots \{1 - \alpha^{-1} \alpha_N(t)\}$$

and we consider the  $t$ -roots of the equation

$$(4.3) \quad |\alpha I - P(t)| = 0.$$

If  $|\alpha| > \alpha_1(u)$  ( $t = u + iv$ ) it follows from (4.2) and Theorem 1(c) that  $|\alpha I - P(t)| \neq 0$ . Thus there can be no  $t$ -roots of (4.3) in any part of the  $t$ -plane for which  $|\alpha| > \alpha_1(u)$ .

Now suppose  $\alpha_1(u)$  is a concave function of  $u$  in some interval  $(u', u'')$ . (The argument will be simpler to follow with the aid of a diagram of the  $u, \alpha_1(u)$  plane). We may choose real numbers  $a$  and  $b$  so that the linear function  $a + bu$



satisfies

$$(4.4) \quad a + bu > \alpha_1(u) \quad (u' < u < u''),$$

i.e., the line  $a + bu$  lies above the curve  $\alpha_1(u)$  in the interval  $(u', u'')$ . In (4.2) let  $\alpha = a + bt$ . Since

$$|a + bt| = \{(a + bu)^2 + b^2v^2\}^{1/2} \geq a + bu > \alpha_1(u) \quad (u' < u < u''),$$

there are no roots of the equation

$$(4.5) \quad |(a + bt)I - P(t)| = 0$$

in the strip of the  $t$ -plane  $u' < u < u''$ . But the  $t$ -roots of (4.5) are continuous functions of  $a$ , and we may choose values of  $a$  and  $b$  so that the line  $a + bu$  cuts the curve  $\alpha_1(u)$  in two points, thus producing two roots of (4.5) in the strip  $(u', u'')$ . Thus for a suitable  $b$ , there is a value of  $a$ , say  $a'$ , such that for  $a > a'$  there are no roots of (4.5) in the strip  $(u', u'')$ , while for  $a < a'$  there are two roots. This contradicts the continuity of the  $t$ -roots of (4.5) and therefore  $\alpha_1(u)$  cannot be concave in any interval.

Further,  $\alpha_1(u)$  cannot be a linear function. For if  $\alpha_1(u) = 1 + cu$  ( $c \neq 0$ ), say, we can choose a real number  $\beta$  so that the function  $e^{\beta u}(1 + cu)$  is concave near the point  $u = 0$ . But  $e^{\beta t}\alpha_1(t)$  is the maximal eigenvalue of the matrix  $e^{\beta t}P(t)$ , which is of the same type as  $P(t)$ , and which cannot therefore have a concave maximal eigenvalue.

It follows that  $\alpha_1(u)$  is strictly convex.

We may specialize our results to integral random variables. To this end, let  $\phi_{jk}(z)$  be a probability generating function associated with the transition  $(j, k)$  and suppose that there is an annulus  $r_0 < |z| < r'_0$  ( $0 \leq r_0 < 1 < r'_0 \leq \infty$ ) in which all the  $\phi_{jk}(z)$  have convergent Laurent series. Let  $Q(z)$  denote the matrix  $\{p_{jk}\phi_{jk}(z)\}$  and we suppose that  $Q(z)$  is not of the degenerate form

$$(4.6) \quad Q(z) = z^\beta Z P Z^{-1},$$

where  $\beta$  is an integer and  $Z$  is a diagonal matrix of integral powers of  $z$ . For real and positive  $z$  let  $a_1(z)$  be the maximal positive eigenvalue of  $Q(z)$ . If we set  $z = e^t$ , then, by Theorem 2,  $a_1(e^t)$  is a strictly convex function of  $t$  ( $t$  real) and therefore  $a_1(z)$ , though not necessarily convex, has the property of not having a local maximum for real positive  $z$ . This generalizes the result of Volkov [4] who demonstrated this property in the special case where  $\phi_{jk}(z) = z^{a_{jk}}$ .

We return to the matrix  $P(t)$  as defined in Section 2. The convexity property of  $\alpha_1(u)$  ( $t = u + iv$ ) raises the question of whether  $\alpha_1(u)$  attains its unique minimum at a finite value of  $u$ . The answer is clearly affirmative if  $\alpha'_1(0) = 0$ . If  $\alpha'_1(0) < 0$  say, then either  $\alpha_1(u)$  continues to decrease as  $u$  increases or it reaches a minimum and then starts increasing. We distinguish between the cases where the strip (2.1) includes the entire right half-plane ( $u'_0 = \infty$ ) and where it is bounded to the right ( $u'_0 < \infty$ ). Modifications for the left half plane will be obvious (i.e., for the case where  $\alpha'_1(0) > 0$ ).

THEOREM 3. Let  $t = u + iv$  and suppose that  $P(t)$  is not degenerate.

(a) Suppose that  $\alpha_1(u)$  is defined for all  $u > 0$  (i.e.,  $u'_0 = \infty$ ). Then a necessary and sufficient condition for  $\alpha_1(u)$  to be uniformly bounded (and so monotonic decreasing) for all  $u > 0$  is that there exists a diagonal matrix  $D(t)$  of degenerate moment generating functions such that each element of the matrix

$$(4.7) \quad Q(t) = \{D(t)\}^{-1}P(t)D(t)$$

is of the form  $p_{jk}q_{jk}(t)$ ,  $q_{jk}(t)$  being the moment generating function of a non-positive random variable. In the case of integral random variables, each element of  $D(t)$  and each  $q_{jk}(t)$  will be the moment generating function of an integral random variable.

(b) Let  $\alpha'_1(0) < 0$  and suppose  $u'_0 < \infty$ . Then  $\alpha_1(u)$  attains its unique stationary minimum at a finite positive value of  $u$  if one of the following conditions is satisfied:

(i) There exists a number  $u_1$  ( $0 < u_1 < u'_0$ ) such that for each  $j, k$  for which  $f_{jk}(t)$  is defined,  $f'_{jk}(u_1) \geq 0$ .

(ii) For some  $j, k$ ,  $f_{jk}(u) \rightarrow \infty$  as  $u \rightarrow u'_0 -$ .

PROOF.

(a) If (4.7) is satisfied, then  $P(t)$  and  $Q(t)$  have the same eigenvalues. Since each element of  $Q(t)$  is non-increasing for  $t > 0$ , it follows from Section 3(c) that  $\alpha_1(t)$  is non-increasing for  $t > 0$  and therefore bounded for all  $t > 0$ .

Conversely, if  $\alpha_1(u)$  is bounded for  $u > 0$ , we note that for each  $j, k$  for which  $p_{jk} > 0$ , and for some finite, real  $\beta_{jk}$ ,  $\Pr(X > \beta_{jk}) = 0$ , where  $X$  is a random variable with moment generating function  $f_{jk}(t)$ . For if not, then we can find  $n$  and  $j$  such that  $\Pr(S_n > 0 | k_n = j, k_n = j) > 0$ , which implies that  $f_{jj}^{(n)}(u) \rightarrow \infty$  as  $u \rightarrow \infty$ . But this contradicts the boundedness of  $\alpha_1(u)$  since  $\{\alpha_1(u)\}^n \geq p_{jj}^{(n)}(u)$ . Thus for each  $j, k$ ,  $f_{jk}(t)$  represents a random variable which is bounded above and we may write

$$(4.8) \quad f_{jk}(t) = \exp(\beta_{jk}t)g_{jk}(t),$$

where  $\beta_{jk}$  is real for each  $j, k$  and

$$(4.9) \quad g_{jk}(t) = o(e^{\epsilon t}) \quad (t \rightarrow +\infty)$$

for every  $\epsilon > 0$ . Let  $\{x_j(t)\}$  be a right eigenvector of  $P(t)$  corresponding to the eigenvalue  $\alpha_1(t)$ . We can choose  $x_j(t)$  to be the co-factor of, say, the element in position  $(1, j)$  of the matrix  $[\alpha_1(t)I - P(t)]$  ( $j = 1, \dots, N$ ). Thus, for each  $j$ ,  $x_j(t)$  is expressible as a sum of products of the elements of  $[\alpha_1(t)I - P(t)]$ . Hence from (4.8), (4.9) and the boundedness of  $\alpha_1(t)$  ( $t > 0$ ), it follows that there is a finite real number  $\beta_j$  such that

$$(4.10) \quad x_j(t) = y_j(t) \exp(\beta_j t) \quad (j = 1, \dots, N),$$

where for each  $j$  and every  $\epsilon > 0$

$$(4.11) \quad y_j(t) = o(e^{\epsilon t}) \quad (t \rightarrow +\infty).$$

Let  $T(t) = \text{diag} \{x_j(t)\}$ . Then the matrix

$$(4.12) \quad \{r_{jk}(t)\} = [\alpha_1(t)]^{-1} [T(t)]^{-1} P(t) T(t)$$

is a stochastic transition matrix for each real  $t$  by Lemma 3.1 and hence for all real  $t$  we have  $0 \leq r_{jk}(t) \leq 1$ . From (4.12) we have for each  $j, k$

$$p_{jk} f_{jk}(t) x_k(t) = x_j(t) \alpha_1(t) r_{jk}(t),$$

and from the relations (4.8) to (4.11) it follows that  $\beta_{jk} + \beta_k \leq \beta_j$  for each  $j, k$  for which  $p_{jk} > 0$ . The result now follows by taking  $D(t) = \text{diag} \{\exp(\beta_j t)\}$ .

In the case where each  $f_{jk}(t)$  is the moment generating function of an integral random variable, each  $\beta_{jk}$  and  $\beta_j$  will be an integer.

(b) In (i), we have  $\alpha_1(u_1) \geq 0$  since  $\alpha_1(u)$  is a nondecreasing function of each of the elements, and thus  $\alpha_1(u)$  must attain its minimum in the interval  $0 < u \leq u_1$ . In (ii) suppose that for some fixed  $j, k$ ,  $f_{jk}(u) \rightarrow \infty$  as  $u \rightarrow u'_0 -$ . We choose  $n$  so that  $P^n > 0$  and since  $u'_0 < \infty$ , we can find  $C > 0$  such that  $f_{kj}^{(n)}(u) \geq C$  as  $u \rightarrow u'_0 -$ . Then we have

$$\begin{aligned} \{\alpha_1(u)\}^{n+1} &\geq f_{jj}^{(n+1)}(u) \geq p_{jk} p_{kj}^{(n)} f_{jk}(u) f_{kj}^{(n)}(u) \\ &\geq C p_{jk} p_{kj}^{(n)} f_{jk}(u) \rightarrow \infty \quad \text{as } u \rightarrow u'_0 -. \end{aligned}$$

Thus  $\alpha_1(u) \rightarrow \infty$  as  $u \rightarrow u'_0 -$  and the result follows.

We now explore further the properties of  $\alpha_1(t)$  in the case of integral random variables. We first state a well known result concerning characteristic functions of integral random variables.

**LEMMA 4.1.** Let  $f(iv) = E(e^{ivX})$  ( $v$  real) where  $X$  is a non-degenerate integral random variable. Then  $|f(iv_1)| = 1$  for some  $v_1$  ( $v_1 \neq 0$ ,  $-\pi \leq v_1 \leq \pi$ ) if and only if  $v_1/2\pi$  is a rational number, say  $v_1 = 2\pi p/q$  (g.c.d.  $[p, q] = 1$ ,  $q > 1$ ), and  $f(iv)$  is of the form  $e^{imv} g(iv)$ , where  $m$  is an integer and  $g(iv)$  is a characteristic function of period  $2\pi/q$  in  $v$ , or equivalently if and only if  $X$  only takes values of the form  $m + nq$  ( $n = 0, \pm 1, \pm 2, \dots$ ;  $m, q$  integral,  $q > 1$ ).

In the following theorem we prove a corresponding result for the functions  $\alpha_j(iv)$  and it is sufficient to suppose that each  $f_{jk}(t)$  exists only on the imaginary axis.

**THEOREM 4.** Let  $t = iv$  ( $v$  real) and suppose that each of the functions  $f_{jk}(iv)$  is a characteristic function of an integral random variable. Let  $\alpha_j(iv)$  ( $j = 1, \dots, N$ ) be the eigenvalues of  $P(iv)$  where  $\alpha_1(0) = 1$ . Then there exists a number  $v_1 \neq 0$  ( $-\pi \leq v_1 \leq \pi$ ) satisfying  $\alpha_j(iv_1) = 1$  for some  $j$  if and only if  $v_1/2\pi$  is a rational number, say  $v_1 = 2\pi p/q$  (g.c.d.  $[p, q] = 1$ ,  $q > 1$ ), and  $P(iv)$  is of the form

$$(4.13) \quad P(iv) = e^{imv} D(iv) Q(iv) \{D(iv)\}^{-1},$$

where

(i)  $Q(iv) = \{p_{jk} g_{jk}(iv)\}$ , each  $g_{jk}$  being a characteristic function of period  $2\pi/q$  in  $v$ , possibly  $g_{jk}(iv) \equiv 1$ ;

(ii)  $D(iv) = \text{diag} \{\exp(im_p v)\}$  ( $m_1, \dots, m_N$  integral);

(iii)  $m$  is integral.

PROOF. If  $P(iv)$  is of the form (4.13) then we may take  $v_1 = 2\pi/q$  and then  $\exp(imv_1)$  will be an eigenvalue of  $P(iv_1)$ .

Conversely, suppose that  $e^{iv}$  is an eigenvalue of  $P(iv_1)$  ( $v_1 \neq 0$ ,  $-\pi \leq v_1 \leq \pi$ ). Since  $\{P(iv_1)\}^* \leq P$  it follows from Section 3(c) that

$$(4.14) \quad P(iv_1) = e^{iv} DPD^{-1},$$

where  $D^* = I$ , and hence that

$$(4.15) \quad |f_{jk}^{(n)}(iv_1)| = 1, \quad \text{each } j, k \text{ and } n \text{ such that } p_{jk}^{(n)} > 0.$$

If  $f_{jk}^{(n)}(iv)$  is degenerate for each  $j, k$  and  $n$  for which  $p_{jk}^{(n)} > 0$ , then  $\{[P(iv)]^n\}^* = P^n$  for all  $v$  and  $n$  and thus  $|\alpha_1(iv)| = 1$  (all  $v$ ). Hence, again by Section 3(c)  $P(iv) = \alpha_1(iv)D(v)P\{D(v)\}^{-1}$  where  $D(v) = \text{diag}\{d_j(v)\}$  ( $|d_j(v)| = 1$   $j = 1, \dots, N$ ). For each  $j, k$  and all sufficiently large  $n$ , therefore,

$$\{a_1(iv)\}^n d_j(v)\{d_k(v)\}^{-1}$$

is a degenerate characteristic function, so that  $\alpha_1(iv)$  must be of the form  $e^{i\beta v}$  ( $\beta$  real and constant). It follows from Theorem 1(b) that  $P(iv)$  is of the form (4.13) with  $Q(iv) \equiv P$ .

Otherwise, for some  $j, k$  and  $n$ ,  $f_{jk}^{(n)}(iv)$  is not degenerate and hence  $v_1 = 2\pi p/q$  (g.c.d.  $[p, q] = 1$ ,  $q > 1$ ) by Lemma 4.1. In virtue of (4.15) we may write

$$(4.16) \quad f_{jk}^{(n)}(iv) = \exp(im_{jk}^{(n)}v)g_{jk}(iv), \quad \text{each } j, k, n \text{ such that } p_{jk}^{(n)} > 0$$

where  $g_{jk}^{(n)}(iv)$  is a characteristic function of period  $2\pi/q$  and  $m_{jk}^{(n)}$  an integer. In (4.14) let  $D = \text{diag}\{\exp(i\beta_j)\}$  ( $\beta_j$  real,  $j = 1, \dots, N$ ). Then (4.16) (with  $v = v_1 = 2\pi p/q$ ) implies that

$$(4.17) \quad 2\pi p m_{jk}^{(n)}/q = n\sigma + \beta_j - \beta_k + 2N_{jk}^{(n)}\pi, \quad N_{jk}^{(n)} \text{ integral,}$$

for each  $j, k$  and  $n$  such that  $p_{jk}^{(n)} > 0$ . By evaluating (4.17) at  $n$  and  $n+1$  (where  $n$  is such that  $P^n > 0$ ) we obtain

$$\sigma = 2\pi p m/q + 2M\pi, \quad m, M \text{ integral,}$$

and (4.17) for  $j, h$  and  $k, h$  gives the result

$$(4.18) \quad \beta_j - \beta_k = (m_{jk}^{(n)} - m_{kh}^{(n)})2\pi p/q - 2(N_{jh}^{(n)} - N_{kh}^{(n)})\pi$$

Since the left hand side of (4.18) is independent of  $h$  we may take

$$\exp\{i(\beta_j - \beta_k)\} = \exp\{i(m_j - m_k)2\pi p/q\}, \quad m_1, \dots, m_N \text{ integral.}$$

Now from (4.17) we see that (writing  $m_{jk} = m_{jk}^{(1)}$ )

$$m_{jk}2\pi p/q = (m + m_j - m_k)2\pi p/q + 2N'_{jk}\pi, \quad N'_{jk} \text{ integral.}$$

Hence  $m_{jk} = m + m_j - m_k + N'_{jk}q/p$ . Thus  $N'_{jk}q/p$  must be an integer and since g.c.d.  $(p, q) = 1$  we must have  $m_{jk} = m + m_j - m_k + qM_{jk}$ ,  $M_{jk}$  integral. From (4.16) with  $n = 1$  it follows that  $P(iv)$  is of the form (4.13).

If  $f(iv)$  is the characteristic function of an integral random variable, we may

say that  $f(iv)$  is expressed in its lowest terms if  $f(iv) = e^{imv}g(iv)$ , where  $g(iv)$  has minimal period  $2\pi/q$  ( $q$  integral,  $q \geq 1$ ) and  $m$  is an integer satisfying  $0 \leq m < q$ . Analogously, in the matrix case we may say that  $P(iv)$  is expressed in its lowest terms if it is written in the form (4.13) where  $Q(iv)$  has minimal period  $2\pi/q$  ( $q \geq 1$ ) and  $0 \leq m < q$ . Hence an alternative statement of Theorem 4 is

**THEOREM 4'.** *If  $P(iv)$  is expressed in its lowest terms in the form (4.13), then  $|\alpha_j(iv)| < 1$  ( $0 < |v| \leq \pi; j = 1, \dots, N$ ) if and only if  $q = 1$ .*

**5. The probability of tail values of the sums  $S_n$ .** We use the notation and definitions of Section 2 and we suppose that each  $f_{jk}(t)$  is an analytic moment generating function of an integral random variable. We suppose also that  $P(iv)$ , when expressed in its lowest terms, satisfies the conditions of Theorem 4', i.e., if  $Q(iv)$  has minimal period  $2\pi/q$  ( $q$  integral,  $q \geq 1$ ) where

$$(5.1) \quad P(t) = e^{mt} [\text{diag} \{ \exp(m_j t) \} ] Q(t) [\text{diag} \{ \exp(m_j t) \} ]^{-1},$$

$m, m_1, \dots, m_N$  integral, then  $q = 1$ . If  $P(iv)$  does not satisfy these conditions, i.e., if  $q > 1$ , then we write  $Q_1(iv) = Q(iv/q)$ . Now  $Q_1(iv)$  has minimal period  $2\pi$ , and it would be sufficient to study  $Q_1$  instead of  $Q$ . Hence it is clearly no loss of generality to suppose that  $q = 1$ . Accordingly, we summarize our assumptions concerning  $P(t)$  as follows:

- (i)  $P(iv)$  has minimal period  $2\pi$  in  $v$ ,
- (5.2) (ii)  $P(t)$  is not reducible to the form (5.1) with  $q > 1$ ,
- (iii)  $P(t)$  is not degenerate.

If  $a$  is any real number, then  $e^{-at}\alpha_1(t)$  is the maximal positive eigenvalue of the matrix  $e^{-at}P(t)$  and is therefore a strictly convex function for real  $t$ . We choose  $a$  so that the matrix  $e^{-at}P(t)$  satisfies one of the conditions of Theorem 3 and also so that  $a > \alpha'_1(0)$ , thus ensuring that  $e^{-at}\alpha_1(t)$  attains its unique minimum at a real, positive, finite value of  $t$ . Let

$$m(a) = \inf_{t>0} e^{-at}\alpha_1(t)$$

and let  $t^*(a)$  satisfy  $m(a) = \exp\{at^*(a)\}\alpha_1(t^*(a))$ . Since  $\alpha'_1(0) < a$  we have  $t^*(a) > 0$  and  $0 < m(a) < 1$ . For brevity we write  $t^* = t^*(a)$ . We now define the matrices

$$\Phi_n(a) = \{p_{jk}^{(n)} \Pr(S_n = na \mid k_0 = j, k_n = k)\}$$

and

$$\Pi_n(a) = \{p_{jk}^{(n)} \Pr(S_n \geq na \mid k_0 = j, k_n = k)\},$$

and our task will be to obtain asymptotic expressions for these as  $n \rightarrow \infty$ . We shall follow closely the methods used by Blackwell and Hodges [1].

The matrix  $e^{-at^*}P(t^*)$  is non-negative, irreducible and primitive, so that it has positive right and left eigenvectors  $x^* = (x_j^*)$ ,  $y^* = (y_j^*)$  respectively such

that  $y^*x^* = 1$ . Let

$$r_{jk} = e^{-at^*} \{m(a)\}^{-1} x_k^* (x_j^*)^{-1} p_{jk} f_{jk}(t^*).$$

Then it follows from Lemma 3.1 that  $R = (r_{jk})$  is the transition matrix of an ergodic Markov chain with no cyclically moving sub-classes. Let  $K_n$  denote the state at time  $n$  in a realization of this chain ( $n = 0, 1, 2, \dots$ ). Let

$$(5.3) \quad \begin{aligned} R(t) &= \{r_{jk} f_{jk}(t + t^*) / f_{jk}(t^*)\} && \text{i.e.,} \\ R(t) &= \{m(a)\}^{-1} e^{-at^*} D^{-1} P(t + t^*) D, \end{aligned}$$

where  $D = \text{diag}(x_j^*)$ . For each  $j, k$  for which  $p_{jk} > 0$ ,  $f_{jk}(t + t^*) / f_{jk}(t^*)$  is the moment generating function of an integral random variable. We define a sequence of random variables  $Y_1, Y_2, \dots$  associated with the Markov chain  $K_0, K_1, K_2, \dots$  in such a way that  $Y_n$  has the moment generating function  $f_{jk}(t + t^*) / f_{jk}(t^*)$  if  $K_{n-1} = j$  and  $K_n = k$ . Thus  $Y_1, Y_2, \dots$  are associated with  $R(t)$  in the same way as  $X_1, X_2, \dots$  are associated with  $P(t)$ .

Let  $R^n = (r_{jk}^{(n)})$  and  $T_n = Y_1 + \dots + Y_n$ . If we raise each side of (5.3) to the power  $n$  and equate coefficients of  $e^{nat}$  (assuming  $na$  to be an integer) we obtain the relation

$$(5.4) \quad \begin{aligned} p_{jk}^{(n)} \Pr(S_n = na \mid k_0 = j, k_n = k) \\ = \{m(a)\}^n x_j^* (x_k^*)^{-1} r_{jk}^{(n)} \Pr(T_n = na \mid K_0 = j, K_n = k) \end{aligned}$$

which corresponds to Theorem 1 of Blackwell and Hodges. Further, for any integer  $s$ , we have

$$(5.5) \quad \begin{aligned} p_{jk}^{(n)} \Pr(S_n = na + s \mid k_0 = j, k_n = k) \\ = \{m(a)\}^n x_j^* (x_k^*)^{-1} e^{-st^*} r_{jk}^{(n)} \Pr(T_n = na + s \mid K_0 = j, K_n = k). \end{aligned}$$

Let  $\beta_1(t) = \alpha_1(t + t^*) / \alpha_1(t^*)$ ,  $\beta_2(t), \dots, \beta_N(t)$  be the eigenvalues of  $R(t)$ . Since  $\beta_1'(0) = a$ , the asymptotic expectation of the increment  $T_n - T_{n-1}$  is  $a$ , whereas that of  $S_n - S_{n-1}$  is  $\alpha_1'(0)$ . Thus we have achieved a shift of expectation similar to that of Blackwell and Hodges and others mentioned in [1].

For each  $j, k$  the possible values of  $Y_1$  are identical to those of  $X_1$  and so  $|\beta_1(iv)| < 1$  ( $0 < |v| \leq \pi$ ) by Theorem 4'. Since  $\beta_1(0) = 1$  and  $[R(iv)]^* \leq R$ , it follows from Section 3(c) that  $|\beta_j(iv)| < 1$  ( $j = 2, 3, \dots, N; -\pi \leq v \leq \pi$ ) and hence by continuity that there exists a number  $\eta$  ( $0 < \eta < 1$ ) such that  $|\beta_j(iv)| \leq \eta$  ( $j = 2, 3, \dots, N; -\pi \leq v \leq \pi$ ). Let  $x(t) = \{x_j(t)\}$  and  $y(t) = \{y_j(t)\}$  be respectively right and left eigenvectors of  $R(t)$  corresponding to the root  $\beta_1(t)$ , chosen so that  $x_j(0) = 1, j = 1, \dots, N$  and

$$\sum_{j=1}^N x_j(t) y_j(t) = 1.$$

It follows from the Jordan canonical form for  $R(t)$  that

$$(5.6) \quad \{R(iv)\}^n = x(iv) y(iv) \{\beta_1(iv)\}^n + O(\eta^n).$$



Let  $\sigma^2 = \beta_1''(0) - a^2 (= \alpha_1''(t^*)/\alpha_1(t^*) - a^2)$  and we have

THEOREM 5. *Provided that  $na$  is a possible value of  $S_n$  the following asymptotic matrix relations hold as  $n \rightarrow \infty$*

- (i)  $\Phi_n(a) = \{m(a)\}^n \{\sigma(2\pi n)^{\frac{1}{2}}\}^{-1} x^* y^* \{I + O(n^{-1})\};$
- (ii)  $\Pi_n(a) = \{m(a)\}^n [\sigma(2\pi n)^{\frac{1}{2}}(1 - e^{-t^*})]^{-1} x^* y^* \{I + O(n^{-1})\}.$

PROOF. It follows from (5.6) and the theory of Fourier series that

$$\begin{aligned} & r_{jk}^{(n)} \Pr(T_n = na \mid K_0 = j, K_n = k) \\ (5.7) \quad &= (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-ianv} \{\beta_1(iv)\}^n x_j(iv) y_k(iv) dv + O(\eta^n) \\ &= (2\pi)^{-1} \int_{-\pi}^{\pi} B_n(v) dv + O(\eta^n), \text{ say.} \end{aligned}$$

Since  $\beta_1'(0) = a$  and  $\beta_1''(0) - a^2 = \sigma^2$ , we may choose  $v_0 (>0)$  so that

$$(5.8) \quad |e^{-ianv} \beta_1(iv)| \leq 1 - \sigma^2 v^2/3 \quad (|v| \leq v_0).$$

We break up the range of integration in (5.7) into the ranges  $|v| \leq n^{-\frac{1}{2}} \log n$ ,  $n^{-\frac{1}{2}} \log n < |v| \leq v_0$ , and  $v_0 < |v| \leq \pi$ . In the first of these ranges we have the expansions

$$\log [\{e^{-ianv} \beta_1(iv)\}^n] = -n\sigma^2 v^2/2 + n \sum_{r=3}^{\infty} b_r v^r$$

and

$$x_j(iv) y_k(iv) = x_k^* y_k^* + \sum_{r=1}^{\infty} c_r v^r \quad (c_r = c_r(j, k), r = 1, 2, \dots)$$

since, in the latter expansion,  $y(0)x(0) = 1$  and therefore, by Lemma 3.1,  $y_k(0) = x_k^* y_k^*$ . Thus on setting  $w = n^{\frac{1}{2}} \sigma v$ , the integrand in (5.7) may be written, for  $|w| \leq \sigma \log n$ ,

$$e^{-\frac{1}{2}w^2} [x_k^* y_k^* + w C_1(w^2) n^{-\frac{1}{2}} + C_2(w^2) n^{-1} + o(n^{-1})]$$

where  $C_1$  and  $C_2$  are polynomials in  $w^2$ , depending on  $j$  and  $k$  but not on  $n$ . Using the result that

$$\int_{-\sigma \log n}^{\sigma \log n} t^p e^{-\frac{1}{2}t^2} dt = 2^{\frac{1}{2}(p+1)} \Gamma(\frac{1}{2}(p+1)) + o(n^{-2})$$

when  $p$  is even and vanishes when  $p$  is odd, we have

$$(2\pi)^{-1} \int_{n^{-\frac{1}{2}} \log n}^{n^{-\frac{1}{2}} \log n} B_n(v) dv = (2\pi n \sigma^2)^{-1} x_k^* y_k^* \{1 + O(n^{-1})\}.$$

In virtue of (5.8) we have

$$\int_{n^{-\frac{1}{2}} \log n < |v| \leq v_0} B_n(v) dv = 0 \left( \int_{n^{-\frac{1}{2}} \log n}^{\infty} \exp(-\frac{1}{2} n \sigma^2 v^2) dv \right)$$

which is  $o(n^{-K})$  for all  $K$ . In the range  $v_0 < |v| \leq \pi$ ,  $|\beta_1(iv)| \leq \rho$ , say, ( $0 < \rho < 1$ ) and so

$$\int_{v_0 < |v| \leq \pi} B_n(v) dv = O(\rho^n).$$

Combining these we find finally

$$r_{jk}^{(n)} \Pr(T_n = na | K_0 = j, K_n = k) = (2\pi n \sigma^2)^{-1} x_k^* y_j^* \{1 + O(n^{-1})\}$$

and the result (i) now follows from (5.4). The result (ii) follows from (5.5) by summing with respect to  $s$  ( $s = 0, 1, 2, \dots$ ).

**6. Acknowledgments.** I should like to thank Dr. P. Whittle for helpful suggestions and to acknowledge a grant from the South African Council for Scientific and Industrial Research.

#### REFERENCES

- [1] DAVID BLACKWELL AND J. L. HODGES, JR., "The probability in the extreme tail of a convolution," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1113-1120.
- [2] SALOMON BOCHNER AND WILLIAM TED MARTIN, *Several Complex Variables*, Princeton University Press, Princeton, N. J., 1948.
- [3] GERARD DEBREU AND I. N. HERSTEIN, "Non-negative square matrices," *Econometrica*, Vol. 21 (1953), pp. 597-607.
- [4] I. S. VOLKOV, "On the distribution of sums of random variables defined on a homogeneous Markov chain with a finite number of states," *Teor. Veroyat. i ee Prim.*, Vol. 3 (1958), pp. 413-429 (in Russian).
- [5] I. S. VOLKOV, "On probabilities for extreme values of sums of random variables defined on a homogeneous Markov chain with a finite number of states," *Teor. Veroyat. i ee Prim.*, Vol. 5 (1960), pp. 338-352 (in Russian).

# LIMIT DISTRIBUTIONS IN THE THEORY OF COUNTERS

BY G. SANKARANARAYANAN

*Annammalai University, India*

**1. Introduction.** Let us suppose that particles that arrive on the counter be randomly spaced on the positive time axis. In an actual counter each particle arriving in the time interval  $(0, \infty)$  independently of others gives rise to an impulse with probability  $p$  or 1 according to whether at this instant there is an impulse present or not. Hence the counter is in one or other of two mutually exclusive states which we denote by  $A$  and  $B$ . The counter is in state  $A$  when no impulse covers the instant and is in state  $B$  otherwise and it assumes the states  $A$  and  $B$  alternatively. A particle striking the counter is recorded if and only if the counter is in state  $A$ . If  $p = 0$ , we get the type I counter and if  $p = 1$ , we get the type II counter.

Let  $t_1, t_2, t_3 \dots$  be the instants at which particles arrive and  $\chi_1, \chi_2, \chi_3, \dots$  be the lengths of successive impulses. Let  $\tau_1, \tau_2, \tau_3 \dots$  be the instants of successive recordings. Let us assume that the time from an arbitrary point in the positive time axis to the time of arrival of the first particle that follows it is a random variable with distribution function  $F(x)$ . Hence the differences  $\{t_{r+1} - t_r\}$ ,  $r = 1, 2, 3, \dots$ , are identically distributed random variables independent of each other with a distribution function  $F(x)$ . Let the time durations of the impulses be independent and identically distributed random variables with the distribution function  $H(x)$ . Let these random variables be independent of the instants of arrivals and of the events of the realizations of the impulses. Let  $\nu_t$  denote the number of registered particles in the time interval  $(0, t)$ . Let the process start in state  $A$ . Let us denote by  $\xi_1, \eta_1, \xi_2, \eta_2, \dots$  the times spent in states  $A$  and  $B$  respectively. Consequently  $\{\xi_n\}$  and  $\{\eta_n\}$  are independent sequences of identically distributed positive random variables. It can be seen that  $\Pr[\xi_n \leq x] = F(x)$ ,  $x > 0$ . Let  $\Pr[\eta_n \leq x] = U(x)$ ,  $x > 0$ . It can also be seen that the instants of transitions  $A \rightarrow B$  coincide with the instants  $\tau_n$ ,  $n = 1, 2, 3, \dots$ . Hence the time differences  $\{\tau_{n+1} - \tau_n\}$  are identically distributed random variables whose distribution function  $G(x)$  is given by

$$(1.1) \quad G(x) = F(x) * U(x) = \int_0^x U(x-y) dF(y).$$

In [8] Takács has shown that  $\nu_t$ , suitably normalized, is asymptotically normal. In [9] Takács has considered the asymptotic behavior of  $\nu_t/t$ . Here he has also applied the law of the iterated logarithm, as stated by P. Hartman and A. Wintner [5], to  $\nu_t$ . In this paper we consider the asymptotic ( $\sim$ ) behaviour of  $\nu_t$  when  $F(x)$  and  $H(x)$  follow the stable laws with suitable characteristic exponents and we show that  $\nu_t$ , suitably normalized, tends to the Mittag-Leffler distribution for all counters of the types detailed above.

Received September 4, 1960; revised April 26, 1961.

## 2. Definitions and notations. Let

$$r(s) = \int_0^{\infty} e^{-sz} dF(x)$$

and

$$w(s) = \int_0^{\infty} e^{-sz} dU(x),$$

so that

$$\gamma(s) = \int_0^{\infty} e^{-sz} dG(x) = r(s) \cdot w(s).$$

Also

$$\Pr[\tau_1 \leq x] = F(x).$$

Let

$$\begin{aligned} W(t, n) &= \Pr[v_t \leq n] \\ (2.1) \quad &= 1 - \Pr[\tau_{n+1} \leq t] \\ &= 1 - F(t) * G_n(t), \end{aligned}$$

where  $G_n(t)$  is the  $n$ -fold convolution of  $G(t)$  with itself and  $G_0(t) = 1$  if  $t \geq 0$  and 0 otherwise. So

$$\Pr[v_t = n] = F(t) * G_{n-1}(t) - F(t) * G_n(t).$$

Let  $m_k(t)$  be the  $k$ th moment of  $v_t$ ,

$$\begin{aligned} (2.2) \quad m_k(t) &= \sum_{n=1}^{\infty} n^k [F(t) * G_{n-1}(t) - F(t) * G_n(t)] \\ &= \sum_{n=0}^{\infty} [(n+1)^k - n^k] \cdot [1 - W(t, n)]. \end{aligned}$$

Also

$$\begin{aligned} (2.3) \quad \int_0^{\infty} e^{-st} dm_k(t) &= s \int_0^{\infty} e^{-st} m_k(t) dt \\ &= s \int_0^{\infty} e^{-st} \sum_{n=0}^{\infty} [(n+1)^k - n^k] \cdot [1 - W(t, n)] dt \\ &= s \sum_{n=0}^{\infty} [(n+1)^k - n^k] \int_0^{\infty} e^{-st} [1 - W(t, n)] dt \\ &= r(s) \sum_{n=0}^{\infty} [(n+1)^k - n^k] [\gamma(s)]^n. \end{aligned}$$

Interchange of summation and integration is valid since all the terms on the

right side of (2.2) are positive. Let  $N(t)$  be the number of particles arriving in the counter in the time interval  $(0, t)$ . Let

$$Q_n(t) = \Pr [N(t) = n].$$

So

$$(2.4) \quad \begin{aligned} \Pr [N(t) \leq n] &= \Pr [t \leq t_{n+1}] \\ &= 1 - F_{n+1}(t), \end{aligned}$$

where  $F_{n+1}(t)$  is the  $n+1$ -fold convolution of  $F(t)$  with itself. So  $\sum_0^n Q_r(t) = 1 - F_{n+1}(t)$  and hence

$$(2.5) \quad Q_n(t) = F_n(t) - F_{n+1}(t).$$

**3. Type II counter.** Consider the type II counter. Let  $P_A(t)$  denote the probability that at the instant  $t$ , the system is in state  $A$ .

$$(3.1) \quad \begin{aligned} P_A(t) &= \Pr(t < \tau_1) + \Pr(\tau_2 - \xi_2 < t < \tau_2) + \Pr(\tau_3 - \xi_3 < t < \tau_3) + \dots \\ &= \Pr(t < \tau_1) + [\Pr(t < \tau_2) - \Pr(t \leq \xi_1 + \eta_1)] \\ &\quad + [\Pr(t < \tau_3) - \Pr(t \leq \xi_1 + \eta_1 + \xi_2 + \eta_2)] + \dots \\ &= \sum_{n=0}^{\infty} G_n(t) - \sum_{n=0}^{\infty} G_n(t) * F(t). \end{aligned}$$

$P_A(t)$  can also be got as follows. If we know that in the time interval  $(0, t)$  exactly  $n$  particles arrive at the counter (the probability of which is  $Q_n(t)$ ), then the occurrence points of these  $n$  events may be regarded as independent uniformly distributed points in time interval  $(0, t)$ . The probability that an impulse started at a random point will end before the instant  $t$  is  $p_t = (1/t) \int_0^t H(x) dx$ . Because of independence the probability that all the  $n$  impulses started at random points will end before the instant  $t$  is  $(p_t)^n$ . So

$$(3.2) \quad \begin{aligned} P_A(t) &= \sum_{n=0}^{\infty} Q_n(t) (p_t)^n \\ &= \sum_{n=0}^{\infty} [F_n(t) - F_{n+1}(t)] (p_t)^n \\ &= 1 - (1 - p_t) \sum_{n=0}^{\infty} F_n(t) (p_t)^{n-1}. \end{aligned}$$

Also using (3.1), we see that

$$(3.3) \quad \begin{aligned} \int_0^{\infty} e^{-st} P_A(t) dt &= \int_0^{\infty} e^{-st} \left[ \sum_{n=0}^{\infty} G_n(t) - \sum_{n=0}^{\infty} G_n(t) * F(t) \right] dt \\ &= (s^{-1}) \{ (1 - \gamma(s))^{-1} - r(s)/(1 - \gamma(s)) \} \\ &= [1 - r(s)]/s[1 - \gamma(s)]. \end{aligned}$$

Hence

$$(3.4) \quad \gamma(s) = 1 - (1 - r(s)) \left[ s \int_0^\infty e^{st} P_A(t) dt \right]^{-1}.$$

**4. A lemma.** We now state a lemma which can be got as the converse of a theorem [[4] Theorem 108, p. 168] by an obvious change of variables.

**LEMMA.** Let  $L(x)$  be such that  $L(cx) \sim L(x)$  for every positive  $c$  (as  $x \rightarrow \infty$ ). Then

$$(4.1) \quad \sum_{n=0}^{\infty} a_n e^{-(n/\lambda)^\theta} \sim \lambda^\theta L(\lambda^\theta) \Gamma(\alpha + 1), \quad \text{as } \lambda \rightarrow \infty, \theta > 0, \alpha > 0,$$

if

$$(4.2) \quad A(n) = a_0 + a_1 + a_2 + \dots + a_n \sim (1/\theta) n^\theta L(n^\theta), \quad \text{as } n \rightarrow \infty.$$

In particular, if  $\theta = 1$  and  $e^{-1/\lambda} = x$ , we get

$$(4.3) \quad \sum_{n=0}^{\infty} a_n x^n \sim (\log 1/x)^{-\alpha} L\{[\log(1/x)]^{-1}\} \Gamma(\alpha + 1), \quad \text{as } x \rightarrow 1,$$

if

$$(4.4) \quad A(n) = a_0 + a_1 + \dots + a_n \sim n^\alpha L(n), \quad \text{as } n \rightarrow \infty.$$

Using this lemma in (2.3), we get

$$\begin{aligned} \int_0^\infty e^{-st} dm_k(t) &= r(s) \sum_{n=0}^{\infty} [(n+1)^k - n^k] (\gamma(s))^n \\ &\sim r(s) [\log \{1/\gamma(s)\}]^{-k} \Gamma(k+1), \quad \text{as } s \rightarrow 0. \end{aligned}$$

**5. Asymptotic behavior of  $P_A(t)$  in a type II counter.** We prove the following theorem.

**THEOREM 5.1.** In a type II counter, if  $F(t)$  has a characteristic function  $\phi(\lambda)$  of form

$$(5.1) \quad \log \phi(\lambda) = c\Gamma(-\alpha)[\cos(\pi\alpha/2) - i(\lambda/|\lambda|) \sin(\pi\alpha/2)]|\lambda|^\alpha$$

with  $c > 0$  and  $0 < \alpha < 1$  and further if  $p_t$  be chosen such that

$$(5.2) \quad 1 - p_t \sim \lambda t^{-\beta} \quad \text{as } t \rightarrow \infty$$

where  $\beta < \alpha$  and  $\lambda$  are constants, then

$$(5.3) \quad P_A(t) \sim (c/\lambda\alpha) t^{-(\alpha-\beta)}, \quad \text{as } t \rightarrow \infty.$$

**REMARK I.** In [3] it is shown that the characteristic function  $\phi(\lambda)$  of the stable law is of the form

$$(5.4) \quad \log \phi(\lambda) = i\gamma t + c\Gamma(-\alpha)[\cos(\pi\alpha/2) + i\beta(\lambda/|\lambda|) \sin(\pi\alpha/2)]|\lambda|^\alpha$$

where  $0 < \alpha < 1$ ,  $c > 0$ ,  $-1 \leq \beta \leq 1$ , and  $\gamma$  is any real number. In particular, if  $\gamma = 0$  and  $\beta = -1$  [[6] p. 255], represents the characteristic function of



a positive random variable. In Theorem 5.1 we have taken a distribution function with this characteristic function. This stable law for  $F(t)$  correspond to well known recurrent processes [10, 11] which describe the arrivals of particles to the counter.

REMARK II. If  $H(x) \sim 1 - \lambda(1 - \beta)x^{-\beta}$ , as  $x \rightarrow \infty$ , then  $p_t$  has the property

$$1 - p_t \sim \lambda t^{-\beta}, \quad \text{as } t \rightarrow \infty.$$

PROOF. In our case [6]  $F(t)$  can be put in the form

$$(5.5) \quad F(t) = 1 - [c/\alpha + \epsilon(t)](1/t^\alpha), \quad 0 < t < \infty.$$

where  $\epsilon(t) \rightarrow 0$  as  $t \rightarrow \infty$ . Also

$$(5.6) \quad F_n(t) = F(t/n^{1/\alpha}) = 1 - [nc/\alpha + n \epsilon(t/n^{1/\alpha})](1/t^\alpha).$$

From (3.3)

$$\begin{aligned} P_A(t) &= 1 - (1 - p_t) \sum_{n=1}^{\infty} F_n(t)(p_t)^{n-1} \\ &= (1 - p_t) \sum_{n=1}^{\infty} [1 - F_n(t)](p_t)^{n-1} \\ &= c(1 - p_t)\alpha^{-1}t^{-\alpha} \sum_{n=1}^{\infty} n(p_t)^{n-1} + (1 - p_t)t^{-\alpha} \sum_{n=1}^{\infty} n \epsilon(t/n^{1/\alpha})(p_t)^{n-1} \\ &= I + II \end{aligned}$$

where

$$\begin{aligned} I &= c(1 - p_t)\alpha^{-1}t^{-\alpha} \sum_{n=1}^{\infty} n(p_t)^{n-1} \\ &= c/\{\alpha t^\alpha(1 - p_t)\}, \end{aligned}$$

and

$$II = (1 - p_t)t^{-\alpha} \sum_{n=1}^{\infty} n \epsilon(t/n^{1/\alpha})(p_t)^{n-1}.$$

Consider the sum

$$(5.7) \quad R(t) = (1 - p_t)^2 \sum_{n=1}^{\infty} n \epsilon(t/n^{1/\alpha})(p_t)^{n-1}.$$

Using the theorem in the appendix, when  $\alpha \leq \frac{1}{2}$

$$\begin{aligned} R(t) &\leq (1 - p_t)^2 \sum_{n=1}^{\infty} n(p_t)^{n-1} \{ K'_1 n^{\alpha/(1-\alpha)} t^{-\alpha/(1-\alpha)} + \\ &\quad K'_2 n^{(1+\alpha)/(1-\alpha)} t^{-\alpha(1+\alpha)/(1-\alpha)} + \\ &\quad K'_3 n^{1/(1-\alpha)} t^{-\alpha/(1-\alpha)} + \\ &\quad K'_4 n t^{-\alpha} + K'_5 n^{(2-\alpha)/(1-\alpha)} t^{-\alpha(2-\alpha)/(1-\alpha)} \} \end{aligned}$$

where  $K'_i, i = 1, 2, 3, 4, 5$  are constants independent of  $t$ . So

$$\begin{aligned}
 R(t) &\leq (1 - p_t)^2 \{ (K'_1/t^{a^2/(1-a)}) \sum_1^\infty n^{1/(1-a)} (p_t)^{n-1} + \\
 &\quad (K'_2/t^{a(1+a)/(1-a)}) \sum_1^\infty n^{2/(1-a)} (p_t)^{n-1} + \\
 &\quad (K'_3/t^{a/(1-a)}) \sum_1^\infty n^{(3-a)/(1-a)} (p_t)^{n-1} + \\
 &\quad (K'_4/t^a) \sum_1^\infty n^2 (p_t)^{n-1} \\
 &\quad (K'_5/t^{a(2-a)/(1-a)}) \sum_1^\infty n^{(3-2a)/(1-a)} (p_t)^{n-1} \} \\
 &\sim (1 - p_t)^2 \{ [K'_1/t^{a^2/(1-a)}] \cdot \{1/(1 - p_t)^{(2-a)/(1-a)}\} + \\
 &\quad \{K'_2/t^{a(1+a)/(1-a)}\} \cdot \{1/(1 - p_t)^{(3-a)/(1-a)}\} + \\
 &\quad \{K'_3/t^{a/(1-a)}\} \cdot \{1/(1 - p_t)^{(3-2a)/(1-a)}\} + \\
 &\quad \{K'_4/t^a\} \cdot \{1/(1 - p_t)^3\} + \\
 &\quad \{K'_5/t^{a(2-a)/(1-a)}\} \cdot \{1/(1 - p_t)^{(4-3a)/(1-a)}\} ], \quad t \rightarrow \infty \\
 &\sim \\
 &\quad \{K'_1/t^{a^2/(1-a)}\} \cdot \{1/(1 - p_t)^{a/(1-a)}\} + \\
 &\quad \{K'_2/t^{a(1+a)/(1-a)}\} \cdot \{1/(1 - p_t)^{(1+a)/(1-a)}\} + \\
 &\quad \{K'_3/t^{a/(1-a)}\} \cdot \{1/(1 - p_t)^{1/(1-a)}\} + \\
 &\quad \{K'_4/t^a\} \cdot \{1/(1 - p_t)\} + \\
 &\quad \{K'_5/t^{a(2-a)/(1-a)}\} \cdot \{1/(1 - p_t)^{(2-a)/(1-a)}\}, \quad t \rightarrow \infty.
 \end{aligned}$$

Since  $1 - p_t \sim \lambda t^{-\beta}, t \rightarrow \infty$ ,

$$\begin{aligned}
 R(t) &\leq K'_1/\{\lambda^{a/(1-a)} t^{a(a-\beta)/(1-a)}\} + \\
 &\quad K'_2/\{\lambda^{(1+a)/(1-a)} t^{(1+a)(a-\beta)/(1-a)}\} + \\
 (5.8) \quad &\quad K'_3/\{\lambda^{1/(1-a)} t^{(a-\beta)/(1-a)}\} + \\
 &\quad K'_4/\{\lambda t^{a-\beta}\} + \\
 &\quad K'_5/\{\lambda^{(2-a)/(1-a)} t^{(2-a)(a-\beta)/(1-a)}\}.
 \end{aligned}$$

Hence when  $\beta < \alpha$ ,

$$R(t) \rightarrow 0, \quad \text{as } t \rightarrow \infty.$$

When  $\alpha = \frac{1}{2}$ , using the theorem in the appendix,

$$\begin{aligned}
R(t) &\leq (1 - p_i)^2 \sum_{n=1}^{\infty} n(p_i)^{n-1} \\
&\quad [K_1''(t/n^2)^{-1} + K_2''(t/n^2)^{-1} + K_3''(t/n^2)^{-1} + K_4''(n/t^1)(\log t - 2 \log n)] \\
&\leq (1 - p_i)^2 \sum_{n=1}^{\infty} n(p_i)^{n-1} \\
&\quad [K_1''(n/t^1) + K_2''(n^3/t^1) + K_3''(n^2/t) + K_4''(n \log t)/t^1 - K_4''(2n \log n)/t^1]
\end{aligned}$$

where  $K_i'', i = 1, 2, 3, 4$  are constants independent of  $t$ . So

$$\begin{aligned}
R(t) &\leq (1 - p_i)^2 [(K_1''/t^1) \sum_1^{\infty} n^2(p_i)^{n-1} + (K_2''/t^1) \sum_1^{\infty} n^4(p_i)^{n-1} + \\
&\quad (K_3''/t) \sum_1^{\infty} n^3(p_i)^{n-1} + (K_4'' \log t)/t^1 \sum_1^{\infty} n^2(p_i)^{n-1} - \\
&\quad (2 K_4''/t^1) \sum_1^{\infty} n^2 \log n(p_i)^{n-1}] \\
&\sim (K_1''/t^1) \cdot \{1/(1 - p_i)\} + (K_2''/t^1) \cdot \{1/(1 - p_i)^3\} + \\
&\quad (K_3''/t) \cdot \{1/(1 - p_i)^3\} + (K_4''(\log t)/t^1) \cdot \{1/(1 - p_i)\} - \\
&\quad (2 K_4''/t^1) \cdot \{1/(1 - p_i)^{1+\epsilon}\}.
\end{aligned}$$

So

$$\begin{aligned}
(5.9) \quad R(t) &\leq K_1''/\{\lambda t^{(1-\beta)}\} + K_2''/\{\lambda^3 t^{3(1-\beta)}\} + K_3''/\{\lambda^2 t^{(1-2\beta)}\} + \\
&\quad K_4''(\log t)/\{\lambda t^{(1-\beta)}\} - 2 K_4''/\{\lambda^{1+\epsilon} t^{1-\beta(1+\epsilon)}\}.
\end{aligned}$$

Here  $\epsilon$  is a small positive number which can be taken as small as we like so that  $\epsilon < (1 - 2\beta)/2\beta$ . So when  $\beta < \frac{1}{2}$ ,  $R(t) \rightarrow 0$  as  $t \rightarrow \infty$ . Hence what ever be the value of  $\alpha$  in  $0 < \alpha < 1$ , when  $\beta < \alpha$ ,  $R(t) \rightarrow 0$  as  $t \rightarrow \infty$ . So when  $\beta < \alpha$ , II in (5.8) is negligible compared to I. Hence

$$\begin{aligned}
(5.10) \quad P_A(t) &\sim \{c/(at^\alpha)\} \cdot \{1/(1 - p_i)\}, & t \rightarrow \infty \\
&\sim \{c/(\lambda\alpha)\} \cdot t^{-(\alpha-\beta)}, & t \rightarrow \infty.
\end{aligned}$$

## 6. Asymptotic behavior of $p_i$ in a type II counter.

THEOREM 6.1. With the same assumptions on  $F(t)$  and  $p_i$  as in Theorem 5.1, we have,

$$(6.1) \quad \lim_{t \rightarrow \infty} \Pr \left[ \frac{p_i}{(1/\lambda) \cdot t^\beta \cdot K(\alpha, \beta)} \leq x \right] = g_\beta(x)$$

where

$$(6.2) \quad g_\beta(x) = [1/(\pi\beta)] \int_0^x \left\{ \sum_{j=1}^{\infty} [(-1)^{j-1}/j!] \sin(\pi\beta j) \Gamma(\beta j + 1) y^{j-1} \right\} dy$$

and

$$(6.3) \quad K(\alpha, \beta) = \Gamma(1 - \alpha + \beta)/\Gamma(1 - \alpha).$$

PROOF. From (5.3), we have,

$$(6.4) \quad P_A(t) = \{c/(\lambda\alpha)\}t^{-(\alpha-\beta)}, \quad t \geq t_0.$$

Also  $P_A(t) = 1$ , when  $t = 0$ . So

$$\begin{aligned} s \int_0^\infty e^{-st} P_A(t) dt &= [(cs)/(\lambda\alpha)] \int_{t_0}^\infty e^{-st} t^{-(\alpha-\beta)} dt + s \int_0^{t_0} e^{-st} P_A(t) dt \\ (6.5) \quad &= [(cs)/(\lambda\alpha)] \int_0^\infty e^{-st} t^{-(\alpha-\beta)} dt \\ &\quad - [(cs)/(\lambda\alpha)] \int_0^{t_0} e^{-st} t^{-(\alpha-\beta)} dt + s \int_0^{t_0} e^{-st} P_A(t) dt. \end{aligned}$$

In (6.5), the second integral is less than  $t_0^{(1-\alpha+\beta)}/(1-\alpha+\beta)$  and the third is less than  $M(1 - e^{-st_0}) s^{-1}$  where  $M$  is a constant. So

$$(6.6) \quad s \int_0^\infty e^{-st} P_A(t) dt \sim \frac{c\Gamma(1 - \alpha + \beta)}{\lambda\alpha s^{-(\alpha-\beta)}}, \quad s \rightarrow 0.$$

It follows from (5.1) that

$$r(s) = \int_0^\infty e^{-st} dF(t) = \exp \left[ -\frac{c\Gamma(1 - \alpha)}{\alpha} s^\alpha \right].$$

Using (3.4),

$$\begin{aligned} \gamma(s) &\sim 1 - \frac{[1 - \{1 - (1/\alpha)c\Gamma(1 - \alpha)s^\alpha\}]}{[c\Gamma(1 - \alpha + \beta)]/[\lambda\alpha s^{-(\alpha-\beta)}]}, \quad s \rightarrow 0 \\ &\sim 1 - \frac{\lambda\Gamma(1 - \alpha)s^\beta}{\Gamma(1 - \alpha + \beta)}, \quad s \rightarrow 0. \end{aligned}$$

Using (4.5) we have

$$\begin{aligned} \int_0^\infty e^{-st} dm_k(t) &\sim \frac{[1 - (1/\alpha)c\Gamma(1 - \alpha)s^\alpha]\Gamma(k+1)}{\left\{ -\log \left[ 1 - \frac{\lambda\Gamma(1 - \alpha)}{\Gamma(1 - \alpha + \beta)} s^\beta \right] \right\}^k}, \quad s \rightarrow 0 \\ (6.8) \quad &\sim \frac{\Gamma(k+1)}{\left[ \frac{\lambda\Gamma(1 - \alpha)s^\beta}{\Gamma(1 - \alpha + \beta)} \right]^k}. \end{aligned}$$

Using Karamata's Tauberian theorem [12], we get

$$\begin{aligned}
 (6.9) \quad m_k(t) &\sim \frac{\Gamma(k+1)t^{\beta}}{\{[\lambda\Gamma(1-\alpha)]/\Gamma(1-\alpha+\beta)\}^k \Gamma[\beta k+1]}, & t \rightarrow \infty \\
 &\sim \frac{\Gamma(k+1)}{\Gamma(\beta k+1)} \left[ \frac{t^{\beta} K(\alpha, \beta)}{\lambda} \right]^k, & t \rightarrow \infty
 \end{aligned}$$

where  $K(\alpha, \beta)$  and  $\beta$  are given by (6.3) and (5.2). That is

$$(6.10) \quad E \left[ \frac{\nu_i}{[t^{\beta} K(\alpha, \beta)]/\lambda} \right]^k \sim \frac{\Gamma(k+1)}{\Gamma(\beta k+1)}, \quad t \rightarrow \infty.$$

Hence

$$(6.11) \quad \lim_{t \rightarrow \infty} \Pr \left[ \frac{\nu_i}{[t^{\beta} K(\alpha, \beta)]/\lambda} \leq x \right] = g\beta(x)$$

where  $g\beta(x)$  is defined by (6.2).

**7. Type I counter.** For a type I counter,  $U(x) = H(x)$ . So  $\gamma(s) = r(s) W'(s)$  where  $W'(s) = \int_0^{\infty} e^{-sx} dH(x)$ . Assuming that  $m = \int_0^{\infty} x dH(x)$  exists and  $F(t)$  has the same form as in Theorem 5.1, we find that

$$\begin{aligned}
 (7.1) \quad \gamma(s) &= r(s) W'(s) \\
 &= \exp [(-c/\alpha)\Gamma(1-\alpha)s^{\alpha}] \cdot \{1 - sm + \dots\} \\
 &\sim 1 - (c/\alpha)\Gamma(1-\alpha)s^{\alpha}, & \text{as } s \rightarrow \infty.
 \end{aligned}$$

Hence in this case,

$$(7.2) \quad \int_0^{\infty} e^{-st} dm_k(t) \sim \frac{\Gamma(k+1)}{[(c/\alpha)\Gamma(1-\alpha)s^{\alpha}]^k}, \quad s \rightarrow 0$$

Using Karamata's theorem [12]

$$(7.3) \quad m_k(t) \sim \frac{\Gamma(k+1)t^{\alpha k}}{[(c/\alpha)\Gamma(1-\alpha)]^k \Gamma(\alpha k+1)}, \quad t \rightarrow \infty.$$

Hence we have

**THEOREM 7.1.** For a type I counter with  $F(t)$  as in Theorem 5.1. and an  $H(x)$  having the first moment,

$$(7.4) \quad \lim_{t \rightarrow \infty} \Pr \left[ \frac{\nu_i}{(\alpha t^{\alpha})/[c\Gamma(1-\alpha)]} \leq x \right] = g_{\alpha}(x)$$

where  $g_{\alpha}(x)$  is defined by (6.2).

**8. Actual counter.** In the actual counter described in Section 1,

$$(8.1) \quad P_{\Lambda}(t) = \sum_{j=0}^{\infty} Q'_j(t) (p_i)^j$$

where

$$(8.2) \quad Q'_j(t) = \sum_{n=0}^{\infty} Q_n(t) \binom{n}{j} p^j q^{n-j}.$$

We can write  $P_A(t)$  in the form

$$\begin{aligned} P_A(t) &= \sum_{n=0}^{\infty} Q_n(t) q^n + \frac{(pp_t)}{1!} \sum_{n=1}^{\infty} Q_n(t) \frac{d}{dq} (q^n) \\ &\quad + \frac{(pp_t)^2}{2!} \sum_{n=2}^{\infty} Q_n(t) \frac{d^2}{dq^2} (q^n) + \dots \\ (8.3) \quad &= \left(1 + \frac{pp_t}{1!} \frac{d}{dq} + \frac{(pp_t)^2}{2!} \frac{d^2}{dq^2} + \dots\right) \left(\sum_{n=0}^{\infty} Q_n(t) q^n\right) \\ &= \left(1 + \frac{pp_t}{1!} \frac{d}{dq} + \frac{(pp_t)^2}{2!} \frac{d^2}{dq^2} + \dots\right) \left(1 - (1-q) \sum_{n=1}^{\infty} F_n(t) q^{n-1}\right) \\ &= 1 - [1 - (q + pp_t)] \sum_{n=1}^{\infty} F_n(t) (q + pp_t)^{n-1}. \end{aligned}$$

Hence we have the following theorem.

**THEOREM 8.1.** *With the same assumptions on  $F(t)$  and  $p_t$  as in Theorem 5.1, in the actual counter,*

$$(8.4) \quad P_A(t) \sim [c/(\alpha\lambda p)] t^{-(\alpha-\beta)}, \quad t \rightarrow \infty.$$

**9. Asymptotic behavior of  $\nu_t$  in the actual counter.** By using the same method as the one used by Takács in [9], we can deduce that for the actual counter,

$$(9.1) \quad \gamma(s) = r(s) \left[ \frac{1}{r'(s)} - \frac{(1 - r'(s))}{r'(s)} \left\{ s \int_0^{\infty} e^{-st} P_A(t) dt \right\}^{-1} \right]$$

where  $r'(s) = (pr(s))/(1 - qr(s))$ , is the Laplace Transform of

$$K(t, q) = (1 - q) \sum_{r=1}^{\infty} F_r(t) q^{r-1}$$

which can be verified to be a distribution. It can also be seen that

$$\sum_{j=0}^m Q'_j(t) = 1 - K_{m+1}(t, q)$$

where  $K_{m+1}(t, q)$  is the  $(m+1)$ th convolution of  $K(t, q)$  with itself. Taking  $F(t)$  and  $p_t$  as in Theorem 5.1, with  $\alpha = \beta$  and  $\alpha\lambda p > qc$ , we find that

$$(9.2) \quad \gamma(s) \sim 1 - [(\alpha\lambda p - qc)/(\alpha p)] \Gamma(1 - \alpha) s^{\alpha}.$$

Hence we have the following theorem.

**THEOREM 9.1.** *With the same assumptions on  $F(t)$  and  $p_t$  as in Theorem 6.1, with  $\alpha = \beta$  and  $\lambda\alpha p > qc$ , in the actual counter,*

$$(9.3) \quad \lim_{t \rightarrow \infty} \Pr \left[ \frac{\nu_t}{(\alpha p t^{\alpha}) / [(\alpha\lambda p - qc) \Gamma(1 - \alpha)]} \leq x \right] = g_{\alpha}(x).$$



**Acknowledgment.** The author wishes to express his gratitude to Prof. V. Ganapathy Iyer for helpful suggestions.

### APPENDIX

Here we give a brief sketch of the proof leading to the results in 5.7. At the outset we state a Lemma (the proof of which is fairly simple) which we use in the proof of the accompanying theorem.

LEMMA. Consider the integral

$$(1) \quad I = \int_0^{\delta} \phi^{-\beta} e^{-\alpha T \phi^{-\gamma}} d\phi, \quad \alpha > 0, \quad \gamma > 0, \quad T > 0.$$

If  $\beta > 1$ ,

$$I = [(\alpha T)^{(1-\beta)/\gamma} / \gamma] \cdot [\Gamma(\beta - 1)/\gamma] + K_1(T, \delta) T],$$

If  $\beta < 1$ ,

$$I = K_2(T, \delta),$$

and if  $\beta = 1$ ,

$$I = (1/\gamma)[\log(1/T) + K_3(T, \delta)],$$

where for all  $T$  in the range  $0 < T < \infty$ ,

$$K_i(T, \delta) < K'_i(\delta), \quad i = 1, 2, 3.$$

Note  $K'_i(\delta)$  depends only on  $\delta$  and not on  $T$ .

We now prove the following theorem concerning stable laws.

THEOREM. Consider a distribution function  $F(t)$  with Laplace transform

$$r(s) = \exp \{[-c\Gamma(1-\alpha)/\alpha]s^\alpha\}, \quad c > 0, \quad 0 < \alpha < 1.$$

Then,  $F(t)$  can be put in the form  $1 - F(t) = t^{-\alpha}[(c/\alpha) + \epsilon(t)]$  where

$$(2) \quad \epsilon(t) \leq [K'_1 t^{-\alpha/(1-\alpha)} + K'_2 t^{-\alpha(1+\alpha)/(1-\alpha)} + K'_3 t^{-\alpha/(1-\alpha)} + K'_4 t^{-\alpha} + K'_5 t^{-\alpha(2-\alpha)/(1-\alpha)}]$$

where  $\alpha > \frac{1}{2}$  or  $\alpha < \frac{1}{2}$ , and

$$(3) \quad \epsilon(t) \leq [K''_1 t^{-1} + K''_2 t^{-1} + K''_3 t^{-1} + K''_4 (\log t) t^{-1}]$$

when  $\alpha = \frac{1}{2}$ ,  $K'_i$  and  $K''_i$ ,  $i = 1, 2, 3, \dots$  are positive constants independent of  $t$ .

PROOF. In [6] Mikusiński has shown that

$$(4) \quad \begin{aligned} g(t) &= (1/2\pi i) \int_{-i\infty}^{i\infty} e^{ts-s^\alpha} ds, \quad (0 < t < \infty; 0 < \alpha < 1) \\ &= (1/\pi)[\alpha/(1-\alpha)](1/t) \int_0^\pi u e^{-u} d\phi, \end{aligned}$$

where

$$u = t^{-\alpha/(1-\alpha)} (\sin \alpha\phi / \sin \phi)^{\alpha/(1-\alpha)} \cdot [\sin [(1-\alpha)\phi] / \sin \phi].$$

Put  $T = t^{-\alpha/(1-\alpha)}$  so that  $t = T^{-(1-\alpha)/\alpha}$  and

$$v_1(\phi) = \frac{(\sin \alpha\phi)^{\alpha/(1-\alpha)} (\sin \{(1-\alpha)\phi\})}{(\sin \phi)^{1/(1-\alpha)}}$$

so that  $u = Tv_1(\phi)$ ,  $v_1(\phi)$  and  $v_1'(\phi)$  are continuous and bounded in  $[0, \pi/2]$  and  $v_1(\phi) > 0$  in  $[0, \pi/2]$ .

Let

$$v_2(\phi) = v_1(\pi - \phi) = \phi^{-1/(1-\alpha)} w(\phi).$$

$w(\phi)$  and  $w'(\phi)$  are bounded and continuous in  $[0, \pi/2]$  and  $w(\phi) > 0$  in  $[0, \pi/2]$ . Let  $G(t) = \int_0^t u e^{-u} d\phi$  so that  $g(t) = \{\alpha/[\pi(1-\alpha)t]\} G(t)$ .

We write  $G(t) = G_1(t) + G_2(t)$  where

$$G_1(t) = T \int_0^{\pi/2} v_1(\phi) e^{-Tv_1(\phi)} d\phi.$$

Using the inequality  $0 < 1 - e^{-x} < x$ ,  $x > 0$ , we have

$$\begin{aligned} G_1(t) &= T \left[ \int_0^{\pi/2} v_1(\phi) d\phi + K_4(T)T \right] \\ &= T[A + K_4(T)T] \end{aligned}$$

where  $A = \int_0^{\pi/2} v_1(\phi) d\phi$  and  $|K_4(T)| < A'$  for all  $T$  in  $0 < T < \infty$ ,  $A$  and  $A'$  being constants independent of  $T$ , and

$$G_2(t) = T \int_0^{\pi/2} v_2(\phi) e^{-Tv_2(\phi)} d\phi.$$

In the following analysis we use frequently functions of the form  $K_i(T, \delta)$ ,  $A_i(T, \delta)$ ,  $B_i(T, \delta)$ ,  $C_i(t, \delta)$ ,  $C_i'(t, \delta)$ ,  $C_i''(t, \delta)$ ,  $D_i(t, \delta)$ ,  $D_i'(t, \delta)$ ,  $D_i''(t, \delta)$ ,  $i = 1, 2, 3, \dots$ . These functions which depend upon  $T$  (or  $t$ ) and  $\delta$  have the property that their absolute value will be less than a constant which is independent of  $T$  (or  $t$ ), but may depend upon  $\delta$  for all values of  $T$  (or  $t$ ) in  $0 < t$  (or  $T$ )  $< \infty$ .

We now write  $G_2(t) = G_3(t) + G_4(t)$  where

$$\begin{aligned} G_3(t) &= T \int_0^{\pi/2} \phi^{-1/(1-\alpha)} w(\phi) e^{-T\phi^{-1/(1-\alpha)} w(\phi)} d\phi \\ &= K_5(T, \delta)T, \end{aligned}$$

and

$$G_4(t) = T \int_0^{\pi/2} \phi^{-1/(1-\alpha)} w(\phi) e^{-T\phi^{-1/(1-\alpha)} w(\phi)} d\phi.$$

We now write  $G_4(t) = G_5(t) + G_6(t) + G_7(t)$  where

$$\begin{aligned} G_5(t) &= T \int_0^{\pi/2} \phi^{-1/(1-\alpha)} w(\phi) e^{-T\phi^{-1/(1-\alpha)} w(\phi)} d\phi \\ &= (1-\alpha)[Tw(o)]^{1-\alpha}[\Gamma(\alpha) + K_6(T, \delta)T] \\ G_6(t) &= Tw(o) \int_0^{\pi/2} [\phi^{-1/(1-\alpha)} e^{-T\phi^{-1/(1-\alpha)} w(o)}] [e^{-T\phi^{-1/(1-\alpha)} \{w(\phi) - w(o)\}} - 1] d\phi. \end{aligned}$$

Using the inequality  $|1 - e^{-x}| \leq xe^{|x|}$ ,  $x > 0$  we get

$$|G_6(t)| = K_7(T, \delta) T^2 \int_0^{\delta} \phi^{1-(2/(1-\alpha))} e^{-T\alpha\phi^{-1/(1-\alpha)}} d\phi.$$

Here  $\alpha$  can be equal to  $w(o)/2$  and  $\delta$  is chosen such that in the interval  $0 < \phi < \delta$ ,

$$[w(o) - K\phi] \geq w(o)/2,$$

$K$  being the lowest upper bound of  $|w'(\phi)|$  in  $0 < \phi < \pi/2$ .

Using the Lemma we have

$$|G_6(t)| = [K_8(T, \delta) + K_9(T, \delta)T]T^{2-2\alpha}$$

$$G_7(t) = T \int_0^{\delta} \phi^{1-(1/(1-\alpha))} w_1(\phi) e^{-T\phi^{-1/(1-\alpha)} w(\phi)} d\phi$$

where  $w_1(\phi) = [w(\phi) - w(o)]/\phi$  is bounded. So

$$G_7(t) = K_{10}(T, \delta) T \int_0^{\delta} \phi^{1-(1/(1-\alpha))} e^{-T\alpha'\phi^{-1/(1-\alpha)}} d\phi$$

where  $\alpha'$  is the L.U.B of  $w(\phi)$  in  $[0, \pi/2]$ , which is positive. Hence

$$G_7(t) = \begin{cases} K_{11}(T, \delta) T^{2-2\alpha} + K_{12}(T, \delta) T^2 & \text{if } 1/(1-\alpha) > 2 \text{ or } \alpha > \frac{1}{2} \\ K_{11}(T, \delta) T & \text{if } 1/(1-\alpha) < 2 \text{ or } \alpha < \frac{1}{2} \\ K_{11}(T, \delta) T [\log(1/T) + K_{12}(T, \delta)] & \text{if } 1/(1-\alpha) = 2 \text{ or } \alpha = \frac{1}{2}. \end{cases}$$

Collecting the various terms we can express  $G(t)$  in the form

$$(5) \quad G(t) = (1-\alpha)\Gamma(\alpha)[w(o)T]^{1-\alpha} + A_1(T, \delta)T + A_2(T, \delta)T^2 + A_3(T, \delta)T^{2-\alpha} + A_4(T, \delta)T^{2-2\alpha} + A_5(T, \delta)T^{3-2\alpha},$$

if  $\alpha > \frac{1}{2}$  or  $\alpha < \frac{1}{2}$

and

$$(6) \quad G(t) = (\frac{1}{2})\Gamma(\frac{1}{2})[w(o)T]^{\frac{1}{2}} + B_1(T, \delta)T + B_2(T, \delta)T^2 + B_3(T, \delta)T^{\frac{1}{2}} + B_4(T, \delta)T \log(1/T),$$

if  $\alpha = \frac{1}{2}$ .

That is

$$(7) \quad G(t) = (1-\alpha)\Gamma(\alpha)[w(o)]^{1-\alpha}t^{-\alpha} + C_1(t, \delta)t^{-\alpha/(1-\alpha)} + C_2(t, \delta)t^{-2\alpha/(1-\alpha)} + C_3(t, \delta)t^{-\alpha(2-\alpha)/(1-\alpha)} + C_4(t, \delta)t^{-\alpha(2-2\alpha)/(1-\alpha)} + C_5(t, \delta)t^{-(3-2\alpha)\alpha/(1-\alpha)}$$

if  $\alpha \leq \frac{1}{2}$ .

And

$$(8) \quad G(t) = (\sqrt{\pi}/2)[w(o)]^{\frac{1}{2}}t^{-\frac{1}{2}} + D_1(t, \delta)t^{-1} + D_2(t, \delta)t^{-2} + D_3(t, \delta)t^{-1} + D_4(t, \delta)(\log t)/t$$

if  $\alpha = \frac{1}{2}$ .

So

$$(9) \quad g(t) = (\alpha/\pi)t^{-\alpha-1}[w(o)]^{1-\alpha}\Gamma(\alpha) + C'_1(t, \delta)t^{-1/(1-\alpha)} + \\ C'_2(t, \delta)t^{-(1+\alpha)/(1-\alpha)} + C'_3(t, \delta)t^{(\alpha^2-\alpha-1)/(1-\alpha)} + \\ C'_4(t, \delta)t^{(2\alpha^2-\alpha-1)/(1-\alpha)} + C'_5(t, \delta)t^{(2\alpha^2-2\alpha-1)/(1-\alpha)} \quad \text{if } \alpha \geq \frac{1}{2}.$$

where

$$C'_i(t, \delta) = (\alpha/[\pi(1-\alpha)])C_i(t, \delta),$$

and

$$(10) \quad g(t) = (\frac{1}{2}\sqrt{\pi})[w(o)]^{1-\alpha}t^{-1} + D'_1(t, \delta)t^{-2} + D'_2(t, \delta)t^{-3} \\ + D'_3(t, \delta)t^{-1} + D'_4(t, \delta)t^{-2} \log t \quad \text{if } \alpha = \frac{1}{2}$$

where

$$D'_i(t, \delta) = (1/\pi)D_i(t, \delta)$$

If  $e^{-s\pi}$  is the Laplace transform of  $g(t)$ , then

$$g(t) = (1/2\pi i) \int_{-i\infty}^{i\infty} e^{ts-s\pi} ds.$$

The frequency function whose Laplace transform is  $e^{-ms\pi}$  is given by

$$(11) \quad f(t) = (1/m^{1/\alpha})g(t/m^{1/\alpha})$$

In our case for the distribution considered in (5)  $m = [c\Gamma(1-\alpha)]/\alpha$ .

We first consider the case when  $\alpha < \frac{1}{2}$  or  $\alpha > \frac{1}{2}$ , ( $0 < \alpha < 1$ ). Using (9) and (11)  $f(t)$  can be put in the form

$$(12) \quad f(t) = ct^{-\alpha-1} + C''_1(t, \delta)t^{-1/(1-\alpha)} + \\ C''_2(t, \delta)t^{-(1+\alpha)/(1-\alpha)} + C''_3(t, \delta)t^{(\alpha^2-\alpha-1)/(1-\alpha)} + \\ C''_4(t, \delta)t^{(2\alpha^2-\alpha-1)/(1-\alpha)} + C''_5(t, \delta)t^{(2\alpha^2-2\alpha-1)/(1-\alpha)}$$

Now

$$1 - F(t) = \int_t^\infty f(t) dt.$$

Using (12), after integration,  $F(t)$  can be put in the form

$$(13) \quad 1 - F(t) = (c/\alpha)t^{-\alpha} + t^{-\alpha}\epsilon(t)$$

where

$$(14) \quad \epsilon(t) \leq K'_1 t^{-\alpha^2/(1-\alpha)} + K'_2 t^{-\alpha(1+\alpha)/(1-\alpha)} + K'_3 t^{-\alpha/(1-\alpha)} + \\ K'_4 t^{-\alpha} + K'_5 t^{-\alpha(2-\alpha)/(1-\alpha)}$$

where  $K'_i$  are constants independent of  $t$  for  $0 < t < \infty$ .

In the same manner, where  $\alpha = \frac{1}{2}$ , using (10) and (11),  $f(t)$  can be put in the form

$$(15) \quad f(t) = ct^{-1} + D_1''(t, \delta)t^{-2} + D_2''(t, \delta)t^{-3} + D_3''(t, \delta)t^{-1} + \\ D_4''(t, \delta)t^{-2}[\log t - (\tfrac{1}{2}) \log m]$$

Using (15), after integration,  $F(t)$  can be put in the form

$$1 - F(t) = 2ct^{-1} + t^{-1}\epsilon(t)$$

where

$$(16) \quad \epsilon(t) < K_1''t^{-1} + K_2''t^{-1} + K_3''t^{-1} + K_4''(\log t)/t$$

where  $K_i''$ ,  $i = 1, 2, 3, 4$  are constants independent of  $t$  in  $0 < t < \infty$ . Hence the theorem.

#### REFERENCES

- [1] G. E. ALBERT AND L. NELSON, "Contributions to the statistical theory of counters data," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 9-22.
- [2] W. FELLER, "Fluctuation theory of recurrent events," *Trans. Amer. Math. Soc.*, Vol. 67 (1949), pp. 98-119.
- [3] B. V. GNEDENKO AND A. N. KOLMOGOROV, *Limit Distributions of Sums of Independent Random Variables*, Addison-Wesley Co., Cambridge, Mass., 1954.
- [4] G. H. HARDY, *Divergent Series*, Oxford University Press, London, 1956.
- [5] P. HARTMAN AND A. WINTNER, "On the law of iterated logarithms," *Amer. J. Math.*, Vol. 63 (1941), pp. 169-176.
- [6] P. LÉVY, *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars, Paris, 1954.
- [7] J. MIKUSIŃSKI (Warszawa), "On the function whose Laplace-transform is  $e^{-s^a}$ ," *Studia Mathematica*, Vol. 18 (1959), pp. 191-198.
- [8] L. TAKÁCS, "On a probability problem arising in the theory of counters," *Proc. Cambr. Philos. Soc.*, Vol. 52 (1956), pp. 488-498.
- [9] L. TAKÁCS, "On a probability problem in the theory of counters," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1257-1263.
- [10] L. TAKÁCS, "On a generalization of the renewal theory," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, Vol. 2, Fasc. 1-2, 1957.
- [11] L. TAKÁCS, "On some probability problem concerning the theory of counters," *Academiae Scientiarum Hungaricae*, Vol. 8, Fasc. 1-2, Budapest, 1957.
- [12] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, Princeton, 1946.

# THE TRANSIENT BEHAVIOR OF A SINGLE SERVER QUEUING PROCESS WITH RECURRENT INPUT AND GAMMA SERVICE TIME<sup>1</sup>

BY LAJOS TAKÁCS

Columbia University

**1. Introduction.** Let us consider the following queuing process: Customers arrive at a counter at the instants  $\tau_0, \tau_1, \dots, \tau_n, \dots$  where the interarrival times  $\theta_n = \tau_{n+1} - \tau_n$  ( $n = 0, 1, \dots; \tau_0 = 0$ ) are identically distributed, mutually independent, positive random variables with distribution function  $P\{\theta_n \leq x\} = F(x)$ . We say that  $\{\tau_n\}$  is a *recurrent process*. The customers will be served by a single server. The server is idle if and only if there is no customer waiting at the counter, otherwise the order of the services is irrelevant. The service times are identically distributed, mutually independent random variables with the distribution function

$$(1) \quad H_m(x) = \begin{cases} 1 - \sum_{j=0}^{m-1} e^{-\mu x} \frac{(\mu x)^j}{j!} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

and independent of  $\{\tau_n\}$ .

We are interested in the investigation of the stochastic behavior of the queue size and the busy period of this process. We shall see, however, that if we know the stochastic behavior of the process defined below, then that of the above process can be deduced immediately.

To define the second process let us suppose that customers arrive at a counter in batches of size  $m$  at the instants  $\tau_0, \tau_1, \dots, \tau_n, \dots$ , where  $\{\tau_n\}$  is the recurrent process defined above. There is a single server. The server is idle if and only if there is no customer waiting at the counter, otherwise the order of the services is irrelevant. The service times are identically distributed, mutually independent random variables with the distribution function

$$(2) \quad H(x) = \begin{cases} 1 - e^{-\mu x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

and independent of  $\{\tau_n\}$ .

Denote by  $\xi(t)$  the queue size at the instant  $t$ , i.e.,  $\xi(t)$  is the number of customers waiting or being served at the instant  $t$ . We say that the system is in state  $E_k$  at the instant  $t$  if  $\xi(t) = k$ . Further define  $\xi_n = \xi(\tau_n - 0)$ , i.e.,  $\xi_n$  is the queue size immediately before the arrival of the  $n$ th batch ( $n = 0, 1, \dots$ ).

Received October 20, 1959; revised May 15, 1961.

<sup>1</sup> An invited paper presented on December 27, 1959 at the Washington meeting of the Institute of Mathematical Statistics. This research was sponsored by the Office of Naval Research under Contract Number Nonr-266 (33), Project Number NR 042-034. Reproduction in whole or in part is permitted for any purpose of the United States Government.



If we identify the arrivals of the batches of size  $m$  with the arrivals of individual customers and the total service time of a batch with the service time of an individual customer then the second process reduces to the first one. For, the distribution function of the total service time of a batch in the second process is equal to  $H_m(x)$ , the  $m$ th iterated convolution of  $H(x)$  with itself.

If we consider the first process then the busy period follows the same probability law as in the second process, but the queue size will change to  $[(\xi(t) + m - 1)/m]$ .

REMARK. If we suppose in particular that the batches will be served in the order of their arrival and if  $\eta(t)$  denotes the virtual waiting time at the instant  $t$ , i.e., the time which the first customer in a batch would wait if the batch joined the queue at the instant  $t$ , then we have

$$(3) \quad \eta(t) = \sum_{i=1}^{\xi(t)} \chi_i,$$

where  $\{\chi_i\}$  is a sequence of identically distributed, independent random variables with distribution function  $H(x)$  and independent of  $\xi(t)$ . In this case the waiting time in the first process follows the same probability law as in the second process.

In what follows we shall consider only the second process and determine the stochastic behavior of the queue size and that of the busy period.

The asymptotic behavior of the queue size and that of the waiting time have been investigated already by F. Pollaczek [4], Chapter 7, D. M. G. Wishart [6], and F. G. Foster [2]. The stochastic law of the busy period has been given by B. W. Conolly [1] and it can be deduced from a general theorem of F. Pollaczek [3].

## 2. An auxiliary theorem. Denote by

$$\varphi(s) = \int_0^\infty e^{-sx} dF(x)$$

the Laplace-Stieltjes transform of  $F(x)$  and let

$$\alpha = \int_0^\infty x dF(x).$$

Throughout this paper we use

LEMMA 1. If (a)  $\Re(s) \geq 0$ ,  $|w| < 1$  or (b)  $\Re(s) > 0$ ,  $|w| \leq 1$  or (c)  $\mu\alpha > m$  and  $\Re(s) \geq 0$ ,  $|w| \leq 1$  then the equation

$$(4) \quad z^m = w\varphi(s + \mu(1 - z))$$

has exactly  $m$  roots  $z = \gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , in the unit circle  $|z| < 1$ . We have

$$(5) \quad \gamma_r(s, w) = \sum_{j=1}^{\infty} \frac{(-\mu\epsilon_r)^j w^{j/m}}{j!} \left( \frac{d^{j-1}[\varphi(s + \mu)]^{j/m}}{ds^{j-1}} \right)$$

where  $\epsilon_r = e^{(2\pi r i)/m}$ ,  $r = 1, 2, \dots, m$ , are the  $m$ th roots of unity.

PROOF. In cases (a) and (b) we have  $|w\varphi(s + \mu(1 - z))| < (1 - \epsilon)^m$  if  $|z| = 1 - \epsilon$  and  $\epsilon$  is a sufficiently small positive number. In case (c) we have  $\varphi(\mu\epsilon) < (1 - \epsilon)^m$  if  $\epsilon$  is a sufficiently small positive number. For, if  $0 \leq \epsilon \leq 1$  then  $\varphi(\mu\epsilon)$  and  $(1 - \epsilon)^m$  are monotone decreasing functions of  $\epsilon$ , they agree at  $\epsilon = 0$  and their right-hand derivatives at  $\epsilon = 0$  are  $-\mu\alpha$  and  $-m$  respectively. Hence  $|w\varphi(s + \mu(1 - z))| \leq \varphi(\mu\epsilon) < (1 - \epsilon)^m$  if  $|z| = 1 - \epsilon$  and  $\epsilon$  is small enough. That is in each of the three cases  $|w\varphi(s + \mu(1 - z))| < (1 - \epsilon)^m$  if  $|z| = 1 - \epsilon$  and  $\epsilon > 0$  is small enough. Thus it follows by Rouché's theorem that (4) has exactly  $m$  roots  $z = \gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , in the circle  $|z| < 1 - \epsilon$  or

$$z = \epsilon_r [w\varphi(s + \mu(1 - z))]^{1/m}, \quad r = 1, 2, \dots, m,$$

has exactly one root  $z = \gamma_r(s, w)$  in the circle  $|z| < 1 - \epsilon$ . The explicit form (5) of  $\gamma_r(s, w)$  can be obtained by Lagrange's expansion. (Cf. e.g., E. T. Whittaker and G. N. Watson [5] p. 132.) This completes the proof of the lemma.

We note that the roots  $z = \gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , of the equation (4) are regular functions of  $s$  and  $w$  and by analytical continuation they can be defined also in case  $\mu\alpha \leq m$  for  $\Re(s) \geq 0$  and  $|w| \leq 1$  without changing (5). We have always  $|\gamma_r(s, w)| \leq 1$ ,  $r = 1, 2, \dots, m$ , if  $\Re(s) \geq 0$  and  $|w| \leq 1$ . Note also that (4) has at most one root (possibly multiple) on the unit circle  $|z| = 1$ , namely  $z = 1$  is a root if  $w\varphi(s) = 1$ . Furthermore  $\gamma_r(s, w) = 0$  if and only if  $w = 0$ . If  $w \neq 0$  then the roots  $\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , are distinct.

We remark further that by forming the Lagrange expansion of  $[\gamma_r(s, w)]^k$ ,  $r = 1, 2, \dots, m$ , we can prove that

$$(6) \quad \sum_{r=1}^m [\gamma_r(s, w)]^k = k \sum_{j \geq k/m} \frac{w^j \mu^{mj-k}}{j(mj-k)!} \int_0^\infty e^{-(\mu+s)x} x^{mj-k} dF_j(x),$$

where  $F_j(x)$  denotes the  $j$ th iterated convolution of  $F(x)$  with itself. By using (6) we can obtain explicit formulas for the probabilities considered in this paper.

Finally, we introduce the following abbreviations:  $\gamma_r(s) = \gamma_r(s, 1)$ ,  $g_r(w) = \gamma_r(0, w)$  and  $\omega_r = \gamma_r(0, 1)$ . They are the roots in  $z$  in the unit circle of the equations  $z^m = \varphi(s + \mu(1 - z))$ ,  $z^m = w\varphi(\mu(1 - z))$  and  $z^m = \varphi(\mu(1 - z))$  respectively.

**3. The transient behavior of  $\{\xi_n\}$ .** Define  $a^+ = \max(a, 0)$ . It is easy to see that

$$(7) \quad \xi_{n+1} = [\xi_n + m - \nu_n]^+,$$

where  $\{\nu_n\}$  is a sequence of identically distributed, mutually independent random variables with the distribution

$$(8) \quad \mathbf{P}\{\nu_n = j\} = \int_0^\infty e^{-\mu x} \frac{(\mu x)^j}{j!} dF(x), \quad j = 0, 1, \dots$$

Accordingly the sequence of random variables  $\{\xi_n\}$  forms a homogeneous Markov chain. We say that the system is in state  $E_k$  at the  $n$ th step if  $\xi_n = k$ .

The higher transition probabilities

$$p_{ik}^{(n)} = P\{\xi_n = k \mid \xi_0 = i\}$$

can be obtained by the following

THEOREM 1. If  $|z| \leq 1$ ,  $|w| < 1$ , and  $|y| < 1$  then we have

$$(9) \quad (1-y)(1-w) \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} p_{ik}^{(n)} y^i z^k w^n = \prod_{r=1}^m \left( \frac{1 - g_r(w)}{1 - z g_r(w)} \right) \\ - \frac{y(1-z)(1-w)}{(1-zy)[y^m - w\varphi(\mu(1-y))]} \prod_{r=1}^m \left( \frac{y - g_r(w)}{1 - z g_r(w)} \right),$$

where  $g_r(w)$ ,  $r = 1, 2, \dots, m$  are the  $m$  roots in  $z$  of the equation

$$(10) \quad z^m = w\varphi(\mu(1-z))$$

in the unit circle  $|z| < 1$ .

Instead of proving this theorem we shall prove the more general Theorem 2 from which Theorem 1 can be deduced as a particular case. Theorem 2 determines the joint distribution of  $\tau_n$  and  $\xi_n$  which we need at the investigation of the stochastic law of the busy period. Theorem 2 can be proved in exactly the same way as the more special Theorem 1.

The joint distribution of the random variables  $\tau_n$  and  $\xi_n$  is determined by the probabilities

$$P_{ik}^{(n)}(x) = P\{\tau_n \leq x, \xi_n = k \mid \xi_0 = i\}$$

and these probabilities can be uniquely determined by the Laplace-Stieltjes transforms

$$\pi_{ik}^{(n)}(s) = \int_0^{\infty} e^{-sx} dP_{ik}^{(n)}(x).$$

These Laplace-Stieltjes transforms are given by

THEOREM 2. If  $\Re(s) \geq 0$ ,  $|z| \leq 1$ ,  $|w| < 1$ , and  $|y| < 1$  then we have

$$(11) \quad (1-y)[1 - w\varphi(s)] \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \pi_{ik}^{(n)}(s) y^i z^k w^n = \prod_{r=1}^m \left( \frac{1 - \gamma_r(s, w)}{1 - z \gamma_r(s, w)} \right) \\ - \frac{y(1-z)[1 - w\varphi(s)]}{(1-zy)[y^m - w\varphi(s + \mu(1-y))]} \prod_{r=1}^m \left( \frac{y - \gamma_r(s, w)}{1 - z \gamma_r(s, w)} \right),$$

where  $\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , are the  $m$  roots in  $z$  of the equation

$$(12) \quad z^m = w\varphi(s + \mu(1-z))$$

in the unit circle  $|z| < 1$ .

PROOF. If  $w = 0$  then the theorem is obviously true, therefore we suppose that  $w \neq 0$ . We shall use only the following theorem of the theory of functions of a complex variable: If  $f(z)$  is regular for all finite values of  $z$  and

$$\lim_{|z| \rightarrow \infty} [f(z)/|z|^k] = 0,$$

then  $f(z)$  is a polynomial of degree  $< k$ . If  $k = 1$  then  $f(z)$  is constant.

Let us introduce the generating function

$$\Pi_i^{(n)}(s, z) = \sum_{j=0}^{\infty} \pi_{ij}^{(n)}(s) z^j$$

which is convergent if  $|z| \leq 1$  and  $\Re(s) \geq 0$ . We shall show that if  $|z| = 1$  then  $\Pi_i^{(n)}(s, z)$ ,  $n = 0, 1, \dots$ , satisfies the following recurrence formula

$$(13) \quad \begin{aligned} \Pi_i^{(n+1)}(s, z) = z^m \varphi \left( s + \mu \left( 1 - \frac{1}{z} \right) \right) \Pi_i^{(n)}(s, z) \\ + \sum_{j=0}^{\infty} C_{ij}^{(n+1)}(s) \left( 1 - \frac{1}{z^j} \right), \end{aligned}$$

where for every  $i$  and  $n$   $\sum_{j=0}^{\infty} |C_{ij}^{(n)}(s)| < 1$ .

We have

$$\Pi_i^{(n)}(s, z) = E\{e^{-sr_n} z^{\xi_n} \mid \xi_0 = i\}$$

and further

$$(14) \quad \tau_{n+1} = \tau_n + \theta_n$$

and

$$(15) \quad \xi_{n+1} = [\xi_n + m - \nu_n]^+$$

where  $\{\theta_n, \nu_n\}$  is a sequence of independent vector random variables with distributions  $P\{\theta_n \leq x\} = F(x)$  and

$$P\{\nu_n = j \mid \theta_n = x\} = e^{-\mu x} [(\mu x)^j / j!], \quad j = 0, 1, \dots$$

By (14) and (15) we obtain (13). The first term on the right hand side of (13) is  $E\{e^{-sr_{n+1}} z^{\xi_{n+1} + m - \nu_n} \mid \xi_0 = i\}$ . To obtain  $E\{e^{-sr_{n+1}} z^{\xi_{n+1}} \mid \xi_0 = i\}$  we have to omit from this the terms corresponding to the values  $\xi_n + m - \nu_n = -1, -2, \dots$  and take into consideration that  $\xi_{n+1} = 0$  if and only if  $\xi_n + m - \nu_n \leq 0$ . Thus we obtain the second term on the right hand side of (13), where

$$C_{ij}^{(n+1)}(s) = P\{\xi_n + m - \nu_n = -j \mid \xi_0 = i\} E\{e^{-sr_{n+1}} \mid \xi_n + m - \nu_n = -j, \xi_0 = i\}.$$

To obtain (13) we also used the relation

$$E\{e^{-sr_n} z^{-\nu_n}\} = \varphi(s + \mu(1 - (1/z)))$$

if  $|z| = 1$ .

Now let  $\Re(s) \geq 0$ ,  $|z| \leq 1$ ,  $|w| < 1$ ,  $|y| < 1$  and define

$$A_i(z, s, w) = \sum_{n=0}^{\infty} \Pi_i^{(n)}(s, z) w^n$$

and

$$A(z, s, w, y) = \sum_{i=0}^{\infty} A_i(z, s, w) y^i.$$

Clearly

$$(16) \quad A(z, s, w, y) = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \pi_{ik}^{(n)}(s) y^i z^k w^n$$

and by definition  $A(z, s, w, y)$  is a regular function of  $z$  if  $|z| \leq 1$ ,  $\Re(s) \geq 0$ ,  $|w| < 1$ , and  $|y| < 1$ . If  $|z| = 1$  then by (13) we have

$$A_i(z, s, w) = \frac{z^i + \sum_{n=1}^{\infty} \sum_{j=0}^{\infty} C_{ij}^{(n)}(s) w^n (1 - (1/z^j))}{1 - wz^m \varphi(s + \mu(1 - (1/z)))}$$

and hence if  $|z| = 1$

$$(17) \quad A(z, s, w, y) = \frac{(1 - zy)^{-1} + \sum_{j=0}^{\infty} C_j(s, w, y) (1 - (1/z^j))}{1 - wz^m \varphi(s + \mu(1 - (1/z)))},$$

where the coefficients

$$C_j(s, w, y) = \sum_{n=1}^{\infty} \sum_{i=0}^{\infty} C_{ij}^{(n)}(s) w^n y^i, \quad j = 0, 1, \dots,$$

satisfy the following condition

$$\sum_{j=0}^{\infty} |C_j(s, w, y)| < |w|/(1 - |w|)(1 - |y|).$$

Now let us define  $A(z, s, w, y)$  also for  $|z| > 1$  by (17) if  $\Re(s) \geq 0$ ,  $|w| < 1$ , and  $|y| < 1$ . Thus  $A(z, s, w, y)$  has singularities only at  $z = 1/y$  and at the zeros of the denominator of (17) outside the unit circle. These zeros evidently agree with the reciprocal values of the roots of (4) inside the unit circle. If we define

$$(18) \quad B(z, s, w, y) = A(z, s, w, y) (1 - zy) \prod_{r=1}^m \left( z - \frac{1}{\gamma_r(s, w)} \right)$$

then  $B(z, s, w, y)$  will be a regular function of  $z$  in the whole complex plane. Since obviously

$$\lim_{|z| \rightarrow \infty} [B(z, s, w, y)/|z|^2] = 0$$

therefore  $B(z, s, w, y)$  is a linear function of  $z$ , that is,

$$(19) \quad B(z, s, w, y) = B_0(s, w, y) + zB_1(s, w, y).$$

$B_0(s, w, y)$  and  $B_1(s, w, y)$  can be determined as follows: We have clearly

$$A(1, s, w, y) = \sum_{i=0}^{\infty} \sum_{n=0}^{\infty} [\varphi(s)]^n y^i w^n = \frac{1}{(1 - y)[1 - w\varphi(s)]}$$

and hence by (18)

$$(20) \quad B(1, s, w, y) = \frac{1}{[1 - w\varphi(s)]} \prod_{r=1}^m \left(1 - \frac{1}{\gamma_r(s, w)}\right).$$

Further by (17)

$$\lim_{s \rightarrow 1/y} (1 - zy)A(z, s, w, y) = y^m/[y^m - w\varphi(s + \mu(1 - y))]$$

and hence by (18)

$$(21) \quad B\left(\frac{1}{y}, s, w, y\right) = \frac{1}{[y^m - w\varphi(s + \mu(1 - y))]} \prod_{r=1}^m \left(1 - \frac{y}{\gamma_r(s, w)}\right).$$

Thus (19) is determined by (20) and (21). Finally  $A(z, s, w, y)$  can be obtained by (18). So we get (11) which was to be proved. It is to be remarked that in the above proof we did not exploit the fact that the roots  $\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , are distinct.

REMARK. If we restrict ourselves to the case  $y = 0$  in proving (11) then we have

$$\lim_{|s| \rightarrow \infty} [B(z, s, w, 0)/|z|] = 0,$$

i.e.,  $B(z, s, w, 0)$  is independent of  $z$  and thus it is determined by (20). In this case we obtain by (18) that

$$(22) \quad [1 - w\varphi(s)] \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \pi_{0k}^{(n)}(s) z^k w^n = \prod_{r=1}^m \left( \frac{1 - \gamma_r(s, w)}{1 - z\gamma_r(s, w)} \right)$$

where  $\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$  are defined in Theorem 2.

To prove Theorem 1 let us note that  $p_{ik}^{(n)} = \pi_{ik}^{(n)}(0)$  and thus if  $s = 0$  in (11) then we get (9). In particular if  $s = 0$  in (22) then we get

$$(23) \quad (1 - w) \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} p_{0k}^{(n)} z^k w^n = \prod_{r=1}^m \left( \frac{1 - g_r(w)}{1 - zg_r(w)} \right)$$

where  $g_r(w)$ ,  $r = 1, 2, \dots, m$  are defined in Theorem 1.

#### 4. The limiting distribution of $\{\xi_n\}$ . Using Theorem 1 we shall prove

THEOREM 3. If  $\mu\alpha > m$  then the limiting probability distribution

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\xi_n = k\} = P_k, \quad k = 0, 1, \dots,$$

exists irrespective of the initial distribution. We have

$$(24) \quad \sum_{k=0}^{\infty} P_k z^k = \prod_{r=1}^m \left( \frac{1 - \omega_r}{1 - z\omega_r} \right)$$

where  $\omega_r$ ,  $r = 1, 2, \dots, m$ , are the  $m$  roots in  $z$  of the equation

$$(25) \quad z^m = \varphi(\mu(1 - z))$$

in the unit circle  $|z| < 1$ .



PROOF. Since  $\{\xi_n\}$  is an irreducible and aperiodic Markov chain, the limit  $\lim_{n \rightarrow \infty} p_{ik}^{(n)} = P_k$  always exists irrespective of  $i$  and either every  $P_k > 0$  and  $\{P_k\}$  is a probability distribution or every  $P_k = 0$ . Let  $i = 0$ . Using (23) by Abel's theorem we get

$$\sum_{k=0}^{\infty} P_k z^k = \lim_{w \rightarrow 1} (1-w) \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} p_{0k}^{(n)} z^k w^n = \prod_{r=1}^m \left( \frac{1 - \omega_r}{1 - z\omega_r} \right).$$

If  $\mu\alpha > m$  then  $\{P_k\}$  is a proper probability distribution, because  $|\omega_r| < 1$ ,  $r = 1, 2, \dots, m$ . If  $\mu\alpha \leq m$  then  $\omega_m = 1$  and therefore  $P_k = 0$  for every  $k$ .

Another consequence of Theorem 1 is

THEOREM 4. Denote by  $f_{00}^{(n)}$  the probability that in the Markov chain  $\{\xi_n\}$  starting from state  $E_0$  the first return occurs at the  $n$ th step. If  $|w| < 1$  then

$$(26) \quad \sum_{n=1}^{\infty} f_{00}^{(n)} w^n = 1 - \frac{1-w}{\prod_{r=1}^m [1 - g_r(w)]}$$

where  $g_r(w)$ ,  $r = 1, 2, \dots, m$ , are the  $m$  roots in  $z$  of the equation

$$(27) \quad z^m = w\varphi(s + \mu(1-z))$$

in the unit circle  $|z| < 1$ .

PROOF. By the theory of Markov chains it follows that

$$\sum_{n=1}^{\infty} f_{00}^{(n)} w^n = \frac{\sum_{n=1}^{\infty} p_{00}^{(n)} w^n}{\sum_{n=0}^{\infty} p_{00}^{(n)} w^n}$$

and (26) can be obtained from (23) with  $z = 0$ .

##### 5. The determination of $F_{ik}^{(n)}(x)$ . Let

$$F_{ik}^{(n)}(x) = \mathbf{P}\{\tau_n \leq x, \xi_n = k, \xi_{n-1} > 0, \dots, \xi_1 > 0 \mid \xi_0 = i\}.$$

The Laplace-Stieltjes transform

$$\Phi_{ik}^{(n)}(s) = \int_0^{\infty} e^{-sx} dF_{ik}^{(n)}(x).$$

is given by

THEOREM 5. If  $\Re(s) \geq 0$  and  $|w| < 1$  then we have

$$(28) \quad \sum_{n=1}^{\infty} \Phi_{ik}^{(n)}(s) w^n = \sum_{n=1}^{\infty} \pi_{ik}^{(n)}(s) w^n - \frac{\left( \sum_{n=1}^{\infty} \pi_{i0}^{(n)}(s) w^n \right) \left( \sum_{n=1}^{\infty} \pi_{0k}^{(n)}(s) w^n \right)}{1 + \sum_{n=1}^{\infty} \pi_{00}^{(n)}(s) w^n}$$

where the expressions on the right hand side are defined by (11).

PROOF. By the theorem of total probability we get

$$P_{ik}^{(n)}(x) = F_{ik}^{(n)}(x) + \sum_{j=1}^{n-1} \int_0^x P_{0k}^{(n-j)}(x-y) dF_{i0}^{(j)}(y)$$

and forming Laplace-Stieltjes transforms we have

$$\pi_{ik}^{(n)}(s) = \Phi_{ik}^{(n)}(s) + \sum_{j=1}^{n-1} \pi_{0k}^{(n-j)}(s) \Phi_{i0}^{(j)}(s).$$

Hence

$$(29) \quad \sum_{n=1}^{\infty} \pi_{ik}^{(n)}(s) w^n = \sum_{n=1}^{\infty} \Phi_{ik}^{(n)}(s) w^n + \left( \sum_{n=1}^{\infty} \pi_{0k}^{(n)}(s) w^n \right) \left( \sum_{n=1}^{\infty} \Phi_{i0}^{(n)}(s) w^n \right).$$

If  $k = 0$  in (29) then we get (28) for  $k = 0$ , whence (28) follows for every  $k$  by (29).

By (28) and (11) we conclude

THEOREM 6. If  $\Re(s) \geq 0$  and  $|w| < 1$  then we have

$$(30) \quad \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} \Phi_{0k}^{(n)}(s) z^k w^n = \prod_{r=1}^m \left( \frac{1}{1 - z\gamma_r(s, w)} \right) - \frac{1 - w\varphi(s)}{\prod_{r=1}^m [1 - \gamma_r(s, w)]}$$

where  $\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , are defined in Lemma 1.

In particular if  $z = 0$  in (30) then we have

$$(31) \quad \sum_{n=1}^{\infty} \Phi_{00}^{(n)}(s) w^n = 1 - \frac{1 - w\varphi(s)}{\prod_{r=1}^m [1 - \gamma_r(s, w)]}.$$

REMARK. If  $F_{00}(x)$  denotes the probability that the distance between two consecutive transitions  $E_0 \rightarrow E_m$  is  $\leq x$ , then we have

$$F_{00}(x) = \sum_{n=1}^{\infty} F_{00}^{(n)}(x),$$

and if  $\Phi_{00}(s)$  denotes its Laplace-Stieltjes transform then by (31) we get

$$(32) \quad \Phi_{00}(s) = 1 - \frac{1 - \varphi(s)}{\prod_{r=1}^m [1 - \gamma_r(s)]},$$

where  $\gamma_r(s) = \gamma_r(s, 1)$ ,  $r = 1, 2, \dots, m$ .

6. The probability law of the busy period. Denote by  $G_n(x)$  the probability that a busy period consists in serving  $n$  batches and its length is  $\leq x$ . Write

$$\Gamma_n(s) = \int_0^{\infty} e^{-sx} dG_n(x)$$

if  $\Re(s) \geq 0$ . We shall prove

THEOREM 7. If  $\Re(s) \geq 0$  and  $|w| < 1$  then we have

$$(33) \quad \sum_{n=1}^{\infty} \Gamma_n(s) w^n = 1 - \frac{1 - w \left( \frac{\mu}{\mu + s} \right)^m}{\prod_{r=1}^m \left( 1 - \frac{\mu}{\mu + s} \gamma_r(s, w) \right)}$$

where  $\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , are the  $m$  roots in  $z$  of (4) in the unit circle  $|z| < 1$ .

PROOF. If  $w = 0$  then (33) is evidently true. Thus we suppose that  $w \neq 0$ . By the theorem of total probability we can write that

$$G_1(x) = \mu \int_0^x [1 - F(y)] e^{-\mu y} \frac{(\mu y)^{m-1}}{(m-1)!} dy,$$

and if  $n = 2, 3, \dots$ , then

$$G_n(x) = \mu \sum_{k=1}^{\infty} \int_0^x F_{0k}^{(n-1)}(x-y) [1 - F(y)] e^{-\mu y} \frac{(\mu y)^{k+m-1}}{(k+m-1)!} dy.$$

Hence

$$\Gamma_1(s) = \mu \int_0^{\infty} e^{-(\mu+s)x} \frac{(\mu x)^{m-1}}{(m-1)!} [1 - F(x)] dx$$

and

$$\Gamma_n(s) = \mu \sum_{k=1}^{\infty} \Phi_{0k}^{(n-1)}(s) \int_0^{\infty} e^{-(\mu+s)x} \frac{(\mu x)^{k+m-1}}{(k+m-1)!} [1 - F(x)] dx$$

if  $n = 2, 3, \dots$ . Forming the generating function of  $\{\Gamma_n(s)\}$  we get

$$(34) \quad \sum_{n=1}^{\infty} \Gamma_n(s) w^n = \mu w \sum_{k=0}^{\infty} C_k(s, w) \int_0^{\infty} e^{-(\mu+s)x} \frac{(\mu x)^k}{k!} [1 - F(x)] dx,$$

where  $C_k(s, w) \equiv 0$  if  $k = 0, 1, \dots, m-2$ ;  $C_{m-1}(s, w) \equiv 1$  and

$$C_{k+m-1}(s, w) = \sum_{n=0}^{\infty} \Phi_{0k}^{(n)}(s) w^n, \quad k = 1, 2, \dots.$$

Thus by (30) we have

$$(35) \quad \sum_{k=0}^{\infty} C_k(s, w) z^k = \frac{1}{z} \prod_{r=1}^m \left( \frac{z}{1 - z \gamma_r(s, w)} \right).$$

For fixed  $s$  and  $w$  write

$$(36) \quad f(z) = \prod_{r=1}^m \left( \frac{z}{1 - z \gamma_r(s, w)} \right).$$

Then

$$C_k(s, w) = \frac{1}{2\pi i} \oint_{|z|=1} \frac{f(z)}{z^{k+2}} dz, \quad k = 0, 1, \dots,$$

and by (34) we get

$$(37) \quad \sum_{n=1}^{\infty} \Gamma_n(s) w^n = \frac{\mu w}{2\pi i} \oint_{|z|=1} f(z) \frac{\left[1 - \varphi\left(s + \mu\left(1 - \frac{1}{z}\right)\right)\right]}{z[(\mu + s)z - \mu]} dz.$$

We can integrate term by term because the series is uniformly convergent on the circle  $|z| = 1$ . Now the integral on the right hand side of (37) can be evaluated as  $-2\pi i$  times the sum of the residues of the integrand at the poles  $z = 1/\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ , outside the unit circle. The residue at  $z = 1/\gamma_r(s, w)$  depends on the value  $\varphi(s + \mu(1 - (1/z)))$  at  $z = 1/\gamma_r(s, w)$ , but if  $z = 1/\gamma_r(s, w)$  then

$$\varphi(s + \mu(1 - (1/z))) = 1/wz^m.$$

Accordingly (37) remains unchanged by the substitution

$$\varphi(s + \mu(1 - (1/z))) = 1/wz^m.$$

Hence

$$(38) \quad \sum_{n=1}^{\infty} \Gamma_n(s) w^n = \frac{\mu}{2\pi i} \oint_{|z|=1} f(z) \frac{(w - (1/z^m))}{z[(\mu + s)z - \mu]} dz.$$

On the other hand this integral can be evaluated as  $2\pi i$  times the sum of the residues of the integrand at the poles  $z = 0$  and  $z = \mu/(\mu + s)$  inside the unit circle. Proceeding in this way, we get

$$\sum_{n=1}^{\infty} \Gamma_n(s) w^n = 1 - \left[ \left( \frac{\mu + s}{\mu} \right)^m - w \right] f\left( \frac{\mu}{\mu + s} \right)$$

where  $f(z)$  is defined by (36). This completes the proof of the theorem.

We remark that the above proof would be also valid in the case of multiple roots  $\gamma_r(s, w)$ ,  $r = 1, 2, \dots, m$ . For, if  $z = \gamma_r(s, w)$  is a root of order  $\nu$  then the residue of the integrand in (37) at  $z = 1/\gamma_r(s, w)$  depends on the values

$$\frac{d^i \varphi(s + \mu(1 - (1/z)))}{dz^i} = \frac{d^i [1/wz^m]}{dz^i}, \quad i = 0, 1, \dots, \nu - 1,$$

at  $z = 1/\gamma_r(s, w)$ .

REMARK. Denote by  $G(x)$  the distribution function of the length of the busy period and let  $\Gamma(s)$  be its Laplace-Stieltjes transform. Evidently  $\Gamma(s) = \sum_{n=1}^{\infty} \Gamma_n(s)$  and therefore if  $w \rightarrow 1$  in (33) we get

$$(39) \quad \Gamma(s) = 1 - \frac{1 - \left( \frac{\mu}{\mu + s} \right)^m}{\prod_{r=1}^m \left( 1 - \frac{\mu}{\mu + s} \gamma_r(s) \right)},$$

where  $\gamma_r(s) = \gamma_r(s, 1)$ ,  $r = 1, 2, \dots, m$ .

The probability that a busy period consists in serving  $n$  batches is  $f_{00}^{(n)} = \Gamma_n(0)$  and therefore by (33)

$$(40) \quad \sum_{n=1}^{\infty} f_{00}^{(n)} w^n = 1 - \frac{(1-w)}{\prod_{r=1}^m [1 - g_r(w)]}$$

in agreement with (26).

THEOREM 8. Denote by  $P_{00}(t)$  the probability that the server is idle at the instant  $t$  given that he was idle at  $t = 0$ . If  $\Re(s) > 0$  then

$$(41) \quad \int_0^{\infty} e^{-st} P_{00}(t) dt = \frac{1}{s} - \frac{\left[1 - \left(\frac{\mu}{\mu + s}\right)^m\right]}{s[1 - \varphi(s)]} \prod_{r=1}^m \left( \frac{1 - \gamma_r(s)}{1 - \frac{\mu}{\mu + s} \gamma_r(s)} \right),$$

where  $\gamma_r(s)$ ,  $r = 1, 2, \dots, m$ , are the  $m$  roots in  $z$  of the equation (4) in the unit circle.

PROOF. Clearly  $P_{00}(t) = P\{\xi(t) = 0 \mid \xi(0) = 0\}$ . Denote by  $M_{00}(t)$  the expectation of the number of transitions  $E_0 \rightarrow E_m$  occurring in the time interval  $[0, t]$ , given that  $\xi(0) = 0$ . Then we can write that

$$(42) \quad P_{00}(t) = 1 - \int_0^t [1 - G(t-x)] dM_{00}(x),$$

where

$$M_{00}(t) = I(t) + F_{00}(t) + F_{00}(t) * F_{00}(t) + \dots$$

and  $I(t) = 1$  if  $t \geq 0$ ,  $I(t) = 0$  if  $t < 0$ . Since

$$\int_0^{\infty} e^{-st} dM_{00}(t) = \frac{1}{1 - \Phi_{00}(s)},$$

we get by (42) that

$$\int_0^{\infty} e^{-st} P_{00}(t) dt = \frac{1}{s} \left[ 1 - \frac{1 - \Gamma(s)}{1 - \Phi_{00}(s)} \right]$$

where  $\Phi_{00}(s)$  is defined by (32) and  $\Gamma(s)$  by (39).

#### REFERENCES

- [1] B. W. CONOLLY, "The busy period in relation to the single-server queueing system with general independent arrivals and Erlangian service-time," *J. Roy. Stat. Soc., Ser. B.*, Vol. 22 (1960), pp. 89-96.
- [2] F. G. FOSTER, "Queues with batch arrivals, I," *Acta Math. Acad. Sci. Hungar.*, Vol. 12 (1961), pp. 1-10.
- [3] F. POLLACZEK, "Sur la répartition des périodes d'occupation ininterrompue d'un guichet," *Comptes Rendus Acad. Sci. Paris*, Vol. 234 (1952), pp. 2042-2044.

- [4] FÉLIX POLLACEK, *Problèmes Stochastiques Posés par le Phénomène de Formation d'une Queue d'Attente à un Guichet et par des Phénomènes Apparentés*, Gauthier-Villars, Paris, 1957.
- [5] E. T. WHITTAKER AND G. N. WATSON, *A Course of Modern Analysis*, Cambridge University Press, 1952.
- [6] DAVID M. G. WISHART, "A queueing system with  $\chi^2$  service-time distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 768-779.



# EFFICIENT ESTIMATION OF A REGRESSION PARAMETER FOR CERTAIN SECOND ORDER PROCESSES<sup>1</sup>

BY CHARLOTTE T. STRIEBEL

*Lockheed Missiles and Space Division, Sunnyvale, California*

**0. Summary.** The problem of estimation of a single regression parameter for a process with fixed known regression function and unknown covariance is attacked using a Hilbert space representation of the process. Some general results are obtained which characterize efficiency classes of covariances—that is, classes for each of which there exists a single estimate that is efficient for all members. These results are applied to both the discrete parameter and the continuous parameter stationary process with rational spectral density. Some special results are also obtained concerning the efficiency of the least square estimate.

**1. Introduction.** Let  $x(t)$  be a second order complex-valued process with mean value function zero and covariance

$$(1.1) \quad E[x(t)\overline{x(s)}] = R(t, s),$$

and suppose that the process

$$(1.2) \quad y(t) = k\varphi(t) + x(t)$$

is observed for the parameter  $t$  in a subset  $C^T$  of the real line. The function  $\varphi(t)$  is known, and the parameter  $k$  is to be estimated. The subsets of interest will be the intervals  $(-\infty < t \leq T)$  and  $(0 \leq t \leq T)$  for the continuous parameter process and the integers  $(t = T, T-1, \dots)$  and  $(t = T, T-1, \dots, 0)$  for the discrete parameter process.

A linear unbiased estimate with finite variance will be represented as a linear functional

$$(1.3) \quad \bar{k}^T = \bar{k}^T[y(t), t \in C^T],$$

which is the limit in quadratic mean of unbiased finite linear combinations of the  $y(t)$  process, that is,

$$(1.4) \quad \sum_{i=1}^{M_m} k_{im}^T y(t_{im}^T) \xrightarrow{\text{q.m.}} \bar{k}^T \quad \text{as } m \rightarrow \infty,$$

where

$$(1.5) \quad t_{im}^T \in C^T$$

and

$$(1.6) \quad \sum_{i=1}^{M_m} k_{im}^T \varphi(t_{im}^T) = 1.$$

Received October 31, 1960; revised May 30, 1961.

<sup>1</sup> This paper was prepared under the sponsorship of the Lockheed General Research Program

The limit  $\bar{k}^T$  of (1.4) is a random variable with finite variance. It can be thought of as an element of the  $L_2$  space over the underlying probability space, it can be made to correspond to an element in the reproducing kernel Hilbert space defined by the kernel  $R(s, t)$  (see Parzen [10]), or a correspondence can be set up with elements of another  $L_2$  space as will be done in Section 2. However, it seems more appropriate to use the notation of a linear functional (1.3), since an estimate must finally be reduced to this form so that it can be applied to elements  $y(t)$  of the sample space. Thus the notation  $\bar{k}^T$  will always refer to a particular sequence of coefficients  $\{k_{im}^T\}$  and time points  $\{t_{im}^T\}$  satisfying (1.4)–(1.6), and the expression  $\bar{k}^T[f(t), t \in C^T]$  will indicate the limit in the topology of the range space of  $f(t)$  of the sums  $\sum_i k_{im}^T f(t_{im}^T)$  provided this limit exists.

Since only linear unbiased estimates will be considered, and the criterion by which an estimate will be judged is its variance, it is clear that only second order properties are involved, so that for these purposes the estimation problem is completely determined by the pair  $(R, \varphi)$ . An estimate  $\bar{k}^T$  is said to be *asymptotically efficient* or simply *efficient* for the problem  $(R, \varphi)$  provided

$$(1.7) \quad E(T) = \frac{\text{variance } \bar{k}^T}{\text{variance } \bar{k}} \rightarrow 1 \quad \text{as } T \rightarrow \infty,$$

where  $\bar{k}^T$  is the minimum variance unbiased estimate of  $k$  for the process (1.2) with  $t \in C^T$ .  $E(T)$  will be called the efficiency for the problem  $(R, \varphi)$ .

Interest in efficient estimates arises from the fact that the "best" estimate  $\bar{k}^T$  may be very inconvenient. This estimate is determined by the linear equation

$$(1.8) \quad \bar{k}^T[R(t, s), t \in C^T] = M^T \overline{\varphi(s)}, s \in C^T,$$

where  $M^T$  is a constant. For many problems of interest, the solution to this equation is difficult to exhibit explicitly, and provided it can be computed at all, it will depend on complete knowledge of  $R(t, s)$ . Thus, if the function  $\varphi(t)$ , which will be called the *regression function*, is known, but information concerning the covariance is limited or can be obtained only at considerable expense, it is desirable to find an estimate that is economical of information concerning  $R(t, s)$  in that it is efficient for as wide a class of covariance functions as possible.

The principal estimate that has been proposed is the least square estimate given, for example, by

$$(1.9) \quad k_L^T = \int_0^T \overline{\varphi(t)} y(t) dt / \int_0^T |\varphi(t)|^2 dt$$

for the case  $C^T = (0 \leq t \leq T)$ . This estimate has the advantages that it is easy to compute and requires no knowledge whatever of the covariance. Previous work on the problem of efficient estimates has been restricted to stationary processes, that is,  $R(t, s) = R(t - s)$ , and has been primarily devoted to determining those combinations  $(R, \varphi)$  for which the least square estimate is efficient.

For the continuous parameter Ornstein-Uhlenbeck process,

$$(1.10) \quad R(\tau) = e^{-\beta|\tau|},$$

and for regression functions

$$(1.11) \quad \varphi(t) = t^r \quad \text{or} \quad e^{i\lambda_0 t},$$

where  $r$  is a non-negative integer and  $\lambda_0$  is a real frequency, Mann and Moranda [9] proved that the least square estimate is efficient. The author in [13] extended this result to include regression functions of the form

$$(1.12) \quad \varphi(t) = t^r e^{i\lambda_0 t}$$

and showed further that for the more general function,

$$(1.13) \quad \varphi(t) = \sum_{\alpha=1}^n \varphi_{\alpha} t^r e^{i\lambda_{\alpha} t},$$

where the  $\varphi_{\alpha}$  are non-zero constants, the  $\lambda_{\alpha}$  are real and distinct, and  $n > 1$ , the least square estimate is not efficient.

For a much broader class of covariance function and essentially the same regression functions, this problem was first discussed by Grenander in [2]. Further work was carried out by Grenander and Rosenblatt in [3] and [4]. Rosenblatt considered some of the same problems in the case of vector-valued time series in [11] and extended his results in [12]. Most of these results, together with some examples, appear in Chapter 7 of [5]. In this work only the discrete parameter case is considered, and the regression functions considered are slightly more general than those of the form (1.13). All restrictions on the class of covariances are imposed on the equivalent class of spectral densities  $f(\lambda)$ , which by assumption exist and satisfy the relation

$$(1.14) \quad R(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda t} f(\lambda) d\lambda$$

for a discrete parameter process and

$$(1.15) \quad R(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda t} f(\lambda) d\lambda$$

for a continuous parameter process. In the discrete parameter case for positive continuous spectral density and "slowly increasing" regression function, a necessary and sufficient condition is given in [5] for the least square estimate to be efficient. The same theorem is obtained in [13] for the continuous parameter Ornstein-Uhlenbeck process and regression function of the form (1.13). Theorem 4 in Section 3 extends this result to the continuous parameter processes with rational spectral density.

In Chapter 1.3 of [6] Grenander and Szegő reproduce a few of the results of [5] using the methods of Toeplitz forms. In Chapter 1.4, under certain regularity conditions on  $f(\lambda)$ , he extends his results to the continuous parameter case for the single example

$$(1.16) \quad \varphi(t) = 1.$$

With the exception of those in [6], all the above-mentioned results are derived for the more general problem

$$(1.17) \quad E[y(t)] = \sum_{i=1}^p k_i \varphi_i(t),$$

where the  $k_i$  are unknown parameters and the  $\varphi_i(t)$  are known functions. For  $p > 1$ , the definition of efficiency used by Mann, Moranda, and Striebel is different from that used by Rosenblatt and Grenander. For the case  $p = 1$ , both agree with definition (1.7) made above. In the present paper only the case  $p = 1$  will be considered though it is believed that the results obtained could be generalized to larger values of  $p$ .

In Section 2, for a rather broad class of processes, necessary and sufficient conditions are given for the existence of an estimate that is efficient for two problems  $(R_1, \varphi)$  and  $(R_2, \varphi)$ . When such an estimate exists, it will be said that  $R_1$  and  $R_2$  are *efficiency equivalent*. In Section 3 these results are applied to the problem of a stationary process with rational spectral density and regression function (1.13) where the  $\lambda_\alpha$  are complex with  $\Re \lambda_\alpha = -a \leq 0$ . Both the continuous and discrete cases are considered.

**2. Efficiency equivalence.** It will be assumed that  $\varphi(t)$  and  $R(t, s)$  can be represented as follows:

$$(2.1) \quad R(t, s) = \int_{\Lambda} \xi(t, \lambda) \overline{\xi(s, \lambda)} dF(\lambda),$$

$$(2.2) \quad \varphi(t) = \int_{\Lambda} \xi(t, \lambda) \overline{\Phi^T(\lambda)} dF(\lambda), \quad t \in C^T,$$

where  $\xi(t, \lambda)$  is a complex-valued measurable function on  $R \times R$ , the set

$$(2.3) \quad \Lambda = \bigcup_{\tau} \bigcup_{t \in C^T} \{\lambda \mid \xi(t, \lambda) \neq 0\}$$

is measurable,  $F$  is a measure on the subspace  $(\Lambda, \mathfrak{B})$  of the reals, and  $\Phi^T(\lambda)$  is in the linear span  $L^T(F)$  of  $\{\xi(t, \lambda), t \in C^T\}$  in the Hilbert space  $L_2(\Lambda, \mathfrak{B}, F)$ . Under these assumptions it follows that to each unbiased linear estimate  $\tilde{k}^T$  with finite variance there corresponds an element  $n^T(\lambda)$  in the subspace  $L^T(F)$  such that

$$(2.4) \quad n^T(\lambda) = \tilde{k}^T[\xi(\lambda, t), t \in C^T],$$

$$(2.5) \quad (\Phi^T, n^T) = 1,$$

and

$$(2.6) \quad \text{variance } \tilde{k}^T = (n^T, n^T).$$

The function  $n^T(\lambda)$  corresponding to  $\tilde{k}^T$  is unique a.s.  $F$ . A minimum variance unbiased estimate  $\hat{k}^T$  exists,

$$(2.7) \quad \frac{\Phi^T(\lambda)}{(\Phi^T, \Phi^T)} = \hat{k}^T[\xi(t, \lambda), t \in C^T]$$

and

$$(2.8) \quad \text{variance } \hat{k}^T = 1/(\Phi^T, \Phi^T).$$

These results are fairly standard and can be obtained for example, from more general results by Parzen [10].

The cases which will be considered in the next section are

$$\xi(t, \lambda) = (2\pi)^{-1} e^{i\lambda t},$$

$\Lambda = [-\pi, \pi]$  for the discrete and  $\Lambda = (-\infty, \infty)$  for the continuous parameter stationary process. The solution  $\Phi^T(\lambda)$  of the equation (2.2) will be found by the Wiener-Hopf technique for  $C^T$  half-infinite.

Let  $F_i$  be measures for which there exists  $\Phi_i^T$  satisfying (2.2). Consider  $n_{ij}^T(\lambda)$  in  $L^T(F_i)$  which corresponds to an unbiased estimate  $\hat{k}_j$  for the problem  $(F_i, \varphi)$ . The following measures can then be defined:

$$(2.9) \quad \mu_i^T(B) = \int_B |\Phi_i^T(\lambda)|^2 dF_i(\lambda) / \int |\Phi_i^T(\lambda)|^2 dF_i(\lambda),$$

$$(2.10) \quad \nu_{ij}^T(B) = \int_B |n_{ij}^T(\lambda)|^2 dF_i(\lambda) / \int |n_{ij}^T(\lambda)|^2 dF_i(\lambda).$$

The first subscript on  $n_{ij}^T$  will be omitted when it is clear what problem  $(F_i, \varphi)$  is intended. The efficiency  $E_{ij}(T)$  for the estimate  $n_j^T(\lambda)$  for the problem  $(F_i, \varphi)$  is given by

$$(2.11) \quad \frac{1}{E_{ij}(T)} = \int |n_j^T(\lambda)|^2 dF_i(\lambda) / \int |\Phi_i^T(\lambda)|^2 dF_i(\lambda).$$

LEMMA 1. If  $n_j^T(\lambda)$  is unbiased and efficient for  $(F_i, \varphi)$ , then

$$|\mu_i^T(B) - \nu_{ij}^T(B)| \rightarrow 0 \quad \text{as} \quad T \rightarrow \infty$$

uniformly for  $B \in \mathcal{B}$ .

PROOF. The subscripts will be omitted in the proof. Let

$$a_T = \left[ \int |n^T(\lambda)|^2 dF(\lambda) \right]^{\frac{1}{2}}, \quad b_T = \left[ \int |\Phi^T(\lambda)|^2 dF(\lambda) \right]^{\frac{1}{2}};$$

then

$$\begin{aligned} |\mu^T(B) - \nu^T(B)| &\leq \int \left| \frac{|\Phi^T(\lambda)|^2}{b_T^2} - \frac{|n^T(\lambda)|^2}{a_T^2} \right| dF(\lambda) \\ &\leq \int \left| \frac{\Phi^T(\lambda)}{b_T} - \frac{n^T(\lambda)}{a_T} \right| \left| \frac{\Phi^T(\lambda)}{b_T} + \frac{n^T(\lambda)}{a_T} \right| dF(\lambda) \\ &\leq \left\{ \int \left| \frac{\Phi^T(\lambda)}{b_T} - \frac{n^T(\lambda)}{a_T} \right|^2 dF(\lambda) \int \left| \frac{\Phi^T(\lambda)}{b_T} + \frac{n^T(\lambda)}{a_T} \right|^2 dF(\lambda) \right\}^{\frac{1}{2}} \\ &= \left\{ \left( \frac{b_T^2}{b_T^2} + \frac{a_T^2}{a_T^2} - \frac{2\Re(\Phi^T, n^T)}{a_T b_T} \right) \left( \frac{b_T^2}{b_T^2} + \frac{a_T^2}{a_T^2} + \frac{2\Re(\Phi^T, n^T)}{a_T b_T} \right) \right\}^{\frac{1}{2}}. \end{aligned}$$

The first inequality takes the absolute value under the integral; the second uses

the elementary inequality

$$| |a|^2 - |b|^2 | \leq |a - b| |a + b|;$$

and the third is the Schwarz inequality. Since the estimate is unbiased,

$$(\Phi^T, n^T) = 1,$$

and, since it is efficient,

$$1/E(T) = (n^T, n^T)(\Phi^T, \Phi^T) = a_T^2 b_T^2 \rightarrow 1.$$

LEMMA 2. Let  $n_0^T$  and  $n_1^T$  be unbiased and efficient for  $(R, \varphi)$ . If  $\nu_{11}^T$  converges weakly to a measure  $N_{11}$ ,

$$\nu_{11}^T \xrightarrow{w} N_{11},$$

then  $\nu_{10}^T$  also converges weakly to that measure,

$$\nu_{10}^T \xrightarrow{w} N_{11}.$$

Complete convergence of  $\nu_{11}^T$

$$\nu_{11}^T \xrightarrow{c} N_{11}$$

implies complete convergence of  $\nu_{10}^T$

$$\nu_{10}^T \xrightarrow{c} N_{11}.$$

The terms weak and strong convergence are according to Loève [8]. This lemma is immediate from Lemma 1.

When it is said that  $k_0^T$  or  $n_0^T$  is an estimate for two measures  $F_1$  and  $F_2$ , the following is intended: there are sequences  $\{k_{im}^T\}$  and  $\{t_{im}^T\}$  satisfying (1.4)–(1.6) where convergence is quadratic mean in (1.4) holds for both  $R_1$  and  $R_2$  or equivalently the sequence of functions

$$(2.12) \quad n_m^T(\lambda) = \sum_i k_{im}^T \xi(t_{im}^T, \lambda)$$

converges to  $n_0^T(\lambda)$  in  $L_2(F_1)$  norm and in  $L_2(F_2)$  norm.

THEOREM 1. (i) Let  $A$  be a countable union of intervals on which  $dF_2(\lambda)/dF_1$  exists and is continuous except for a countable number of discontinuities. Consider a sequence  $T$  for which the following are satisfied. (ii) There exist estimates  $n_i^T(\lambda)$  unbiased and efficient for  $(F_i, \varphi)$   $i = 1, 2$  for which

$$\nu_{ii}^T \xrightarrow{w} N_{ii}, \quad i = 1, 2 \quad \text{and} \quad N_{22}(A) \neq 0.$$

Then if there exists an estimate  $n_0^T(\lambda)$  that is unbiased and efficient for  $F_1$  and  $F_2$ , it follows that  $N_{11}$  and  $N_{22}$  must satisfy the following condition: (iii) For all  $B \in \mathcal{B}$

$$(2.13) \quad \int_{B \cap A} \frac{dF_2(\lambda)}{dF_1} dN_{11}(\lambda) = c N_{22}(B \cap A)$$



where

$$(2.14) \quad c = \lim_{T \rightarrow \infty} \frac{\int |n_2^T(\lambda)|^2 dF_2(\lambda)}{\int |n_1^T(\lambda)|^2 dF_1(\lambda)}.$$

PROOF. Let  $(a, b) = B^*$  be an interval contained in  $A$  on which  $dF_2(\lambda)/dF_1$  is continuous, then

$$\int_{B^*} \frac{dF_2(\lambda)}{dF_1} dv_{10}^T(\lambda) = c(T) v_{10}^T(B^*),$$

where

$$c(T) = \frac{\int |n_0^T(\lambda)|^2 dF_2}{\int |n_0^T(\lambda)|^2 dF_1} = \frac{E_{10}(T)E_{22}(T)}{E_{20}(T)E_{11}(T)} \frac{\int |n_2^T(\lambda)|^2 dF_2(\lambda)}{\int |n_1^T(\lambda)|^2 dF_1(\lambda)}.$$

From Lemma 2 and (ii), since  $n_0^T$  is also efficient for  $F_i$ ,

$$v_{i0}^T \xrightarrow{w} N_{ii}, \quad i = 1, 2.$$

By the Helly-Bray Lemma

$$\int_{B^*} \frac{dF_2(\lambda)}{dF_1} dv_{10}^T(\lambda) \rightarrow \int_{B^*} \frac{dF_2(\lambda)}{dF_1} dN_{11}(\lambda).$$

Since  $N_{22}(A) \neq 0$ , there exists an interval  $B^*$  in  $A$  such that  $N_{22}(B^*) \neq 0$ , thus

$$\frac{\int |n_2^T(\lambda)|^2 dF_1}{\int |n_1^T(\lambda)|^2 dF_2} = \frac{c(T)E_{10}(T)E_{22}(T)}{E_{20}(T)E_{11}(T)} \rightarrow c = \frac{\int_{B^*} \frac{dF_2(\lambda)}{dF_1} dN_{11}(\lambda)}{N_{22}(B^*)}.$$

The measurable sets in  $A$  are generated by intervals of this type, so (2.13) must also hold for all  $B \subset \mathcal{B}$ .

THEOREM 2. If in addition to assumptions (i)–(iii),  $c \neq 0$ ,  $A = \Lambda$ ,  $dF_2/dF_1$  is bounded and

$$v_{11}^T \rightarrow N_{11},$$

then  $n_0^T(\lambda)$  efficient and unbiased for  $F_1$  implies  $n_0^T(\lambda)$  is also an efficient unbiased estimate for  $F_2$ .

PROOF. Let  $\{\tilde{k}_{im}^T\}_0, \{t_{im}^T\}_0$  be a sequence of simple estimates (1.4) which converges to  $\tilde{k}_0^T$  in quadratic mean, with respect to  $F_1$ . Then for the corresponding  $\{n_m^T(\lambda)\}$  given by (2.12)

$$\begin{aligned} \int |n_0^T(\lambda) - n_m^T(\lambda)|^2 dF_2 &= \int \frac{dF_2(\lambda)}{dF_1} |n_0^T(\lambda) - n_m^T(\lambda)|^2 dF_1(\lambda) \\ &\leq M \int |n_0^T(\lambda) - n_m^T(\lambda)|^2 dF_1(\lambda) \rightarrow 0, \end{aligned}$$

where  $M$  is the bound of  $dF_2/dF_1$ . Thus (1.4) also converges to an estimate which corresponds to  $n_0^T(\lambda)$  with respect to  $F_2$ .

$$\begin{aligned} \frac{1}{E_{20}(T)} &= \frac{E_{11}(T)}{E_{10}(T)E_{22}(T)} \frac{\int |n_1^T(\lambda)|^2 dF_1}{\int |n_2^T(\lambda)|^2 dF_2} \int \frac{dF_2}{dF_1} d\nu_{10} \\ &\rightarrow \frac{1}{c} \int \frac{dF_2(\lambda)}{dF_1} dN_{11}(\lambda) = N_{22}(\Lambda) \leq 1. \end{aligned}$$

This depends on Lemma 2 and the Helly-Bray Theorem in Section 11.3 of [8].

**3. Rational spectral density and regression function.** In this section the discrete and the continuous parameter stationary process will be considered. Thus

$$(3.1) \quad \xi(t, \lambda) = (2\pi)^{-1} e^{i\lambda t},$$

$$(3.2) \quad \Lambda = [-\pi, \pi]$$

for the discrete parameter process, and

$$(3.3) \quad \Lambda = (-\infty, \infty)$$

for the continuous parameter process, and the representation (2.1) is given by (1.14) and (1.15), respectively. The case of  $C^T$  half-infinite will be considered first. For the discrete parameter  $t$  and  $T$  are integers, and

$$(3.4) \quad C^T = (T, T-1, \dots);$$

for the continuous parameter

$$(3.5) \quad C^T = (-\infty, T).$$

It will be assumed that the spectral densities  $f(z)$  and  $f(\lambda)$  are positive rational functions where for convenience in the discrete parameter case the density will be treated as a function of  $z = e^{i\lambda}$ . The densities can be factored

$$(3.6) \quad f(z) = |F(z)|^2,$$

$$(3.7) \quad f(\lambda) = |F(\lambda)|^2.$$

For the discrete process  $F(z)$  is a quotient of two polynomials each of the same degree and having zeros inside the unit circle ( $|z| < 1$ ); for the continuous process  $F(\lambda)$  is a proper rational function and has poles and zeros in the upper half-plane ( $\Im \lambda > 0$ ). (See Doob [1], p. 502 and p. 542.)

The regression function that will be considered has the form

$$(3.8) \quad \varphi(t) = \sum_{\gamma=1}^m \varphi_\gamma t^\gamma e^{i\lambda_\gamma t}, \quad t \geq 0$$

where  $\lambda_\gamma$  is complex and

$$(3.9) \quad \max_\gamma \Re(i\lambda_\gamma) = a \geq 0 \quad \text{and} \quad \varphi_\gamma \neq 0.$$

The exact form of  $\varphi(t)$  for  $t < 0$  will be seen to be immaterial for questions of efficiency as  $T \rightarrow \infty$ . For the discrete parameter case  $\varphi(t)$  for  $t < 0$  must be such that the sum

$$(3.10) \quad \Phi(z) = \sum_{t=-\infty}^{\infty} z^{-t} \varphi(t)$$

converges to a rational function in a ring

$$(3.11) \quad a < |z| < b.$$

Similarly, in the continuous case the integral

$$(3.12) \quad \Phi(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda t} \varphi(t) dt$$

must converge to a rational function in the strip

$$(3.13) \quad -b < \Im \lambda < -a.$$

In this case it will also be assumed that  $\Phi(\lambda)/F(\lambda)$  is a proper rational function. For any given degree  $e$  and  $\varphi(t)$  given by (3.8) for  $t \geq 0$  it is always possible to define  $\varphi(t)$  for  $t < 0$  so that the degree of denominator of  $\Phi(\lambda)$  exceeds that of the numerator by  $e$  and hence  $\Phi(\lambda)/F(\lambda)$  is proper if the net degree of  $1/F(\lambda)$  is less than  $e$ . In each case the terms of importance in (3.8) are those for which  $\Im(i\lambda_\gamma) = a$  and among these the ones for which  $r_\gamma$  is a maximum. The index of these terms will be indicated by  $\alpha = 1, \dots, n$ . The functions  $\Phi(z)$  and  $\Phi(\lambda)$  can then be expanded as follows:

$$(3.14) \quad \Phi(z) = \sum_{\alpha=1}^n \sum_{j=0}^r \frac{\Phi_{\alpha j} z^{j+1}}{(z - z_\alpha)^{j+1}},$$

where

$$(3.15) \quad \begin{aligned} z_\alpha &= e^{i\lambda_\alpha} \\ |z_\alpha| &= a, & \alpha &= 1, \dots, n, \\ |z_\alpha| &< a \quad \text{or} \quad \geq b, & \alpha &= n+1, \dots, m, \end{aligned}$$

and

$$(3.16) \quad \Phi_{\alpha n} = \Phi_\alpha = r! \varphi_\alpha;$$

$$(3.17) \quad \Phi(\lambda) = \sum_{\alpha=1}^n \sum_{j=0}^r \frac{\Phi_{\alpha j}}{(\lambda - \lambda_\alpha)^{j+1}},$$

$$(3.18) \quad \begin{aligned} \Im \lambda_\alpha &= -a, & \alpha &= 1, \dots, n, \\ \Im \lambda_\alpha &> -a \quad \text{or} \quad \leq -b, & \alpha &= n+1, \dots, m, \end{aligned}$$

$$(3.19) \quad \Phi_{\alpha r} = \Phi_\alpha = r! \varphi_\alpha (-i)^{r+1}.$$

Equation (2.2) can be written

$$(3.20) \quad \frac{1}{(2\pi)^{1/2}} \oint_{|z|=1} z^{t-1} \Phi^T(z) |F(z)|^2 dz = \varphi(t), \quad t = T, T-1, \dots,$$

where

$$(3.21) \quad \overline{\Phi^T(z)} = \sum_{t=-\infty}^T z^{-t} k_t,$$

and

$$(3.22) \quad (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{i\lambda t} \overline{\Phi^T(\lambda)} |F(\lambda)|^2 d\lambda = \varphi(t), \quad -\infty < t \leq T,$$

where

$$(3.23) \quad \overline{\Phi^T(\lambda)} = k^T [e^{-i\lambda t}, -\infty < t \leq T].$$

Under the assumptions made these equations can easily be solved by the Wiener-Hopf technique. (See, for example, [14] p. 313.) Solutions are given by

$$(3.24) \quad \overline{\Phi^T(z)} = \frac{1}{i(2\pi)^{1/2} F(z)} \sum_{t=-\infty}^T z^{-t} \oint_{|w|=e^c} w^{t-1} \frac{\Phi(w)}{F(w)} dw$$

and

$$(3.25) \quad \overline{\Phi^T(\lambda)} = \frac{1}{(2\pi)^{1/2} F(\lambda)} \int_{-\infty}^T e^{-i\lambda t} \int_{-\infty-ic}^{\infty-ic} e^{i\omega t} \frac{\Phi(\omega)}{F(\omega)} d\omega dt,$$

where

$$(3.26) \quad 0 \leq a < c < b.$$

Equation (3.23) can be written as an integral

$$(3.27) \quad \overline{\Phi^T(\lambda)} = \int_{-\infty}^T e^{-i\lambda t} K^T(t) dt$$

if  $K^T(t)$  is permitted to include delta functions and their derivatives. Formulas for the estimates themselves will be given later.

For the case of  $C^T$  half-infinite the "best" estimate  $\hat{k}^T$ , which is clearly efficient, will be considered. For this estimate  $n^T(\lambda)$  is given by (2.7). For a given spectrum  $f_i$ , the measure  $N_{ii}$  and an asymptotic expression for  $\int |n_{ii}^T(\lambda)|^2 f_i(\lambda) d\lambda$  must be obtained in order to apply the theorems of the previous section. These can be obtained by a straightforward but somewhat lengthy calculation and will be given without proof.

LEMMA 3. For the discrete case,  $a > 0$  and a sequence  $T \rightarrow \infty$  such that  $e^{i\pi R \lambda_n} \rightarrow l_\alpha$ ,

$$(3.28) \quad \frac{(\Phi^T, \Phi^T)}{e^{2aT} T^{2r}} \rightarrow c(F) = \sum_{\alpha=1}^n \sum_{\beta=1}^n \frac{\varphi_\alpha \bar{\varphi}_\beta l_\alpha l_\beta z_\alpha \bar{z}_\beta}{F(z_\alpha) \bar{F}(\bar{z}_\beta) (z_\alpha \bar{z}_\beta - 1)} > 0,$$

and

$$(3.29) \quad \frac{|\Phi^T(z)F(z)|^2}{(\Phi^T, \Phi^T)} \xrightarrow{e} \frac{1}{(2\pi)^{1/2}c(F)} \left| \sum_{\alpha=1}^n \frac{\varphi_\alpha l_\alpha z_\alpha}{F(z_\alpha)(z_\alpha - z)} \right|^2.$$

For the discrete case,  $a = 0$  and all sequences  $T \rightarrow \infty$

$$(3.30) \quad \frac{(\Phi^T, \Phi^T)}{T^{2r+1}} \rightarrow c(F) = \frac{1}{2r+1} \sum_{\alpha=1}^n \left| \frac{\varphi_\alpha}{F(z_\alpha)} \right|^2 > 0,$$

and

$$(3.31) \quad \frac{1}{2\pi iz} \frac{|\Phi^T(z)F(z)|^2}{(\Phi^T, \Phi^T)} \xrightarrow{e} \frac{1}{c(F)(2r+1)} \sum_{\alpha=1}^n \left| \frac{\varphi_\alpha}{F(z_\alpha)} \right|^2 \delta(z - a_\alpha)$$

where  $\delta$  is the Dirac delta function. For the continuous parameter case,  $a > 0$  and a sequence  $T \rightarrow \infty$  for which  $e^{iT\Omega a} \rightarrow l_a$

$$(3.32) \quad \frac{(\Phi^T, \Phi^T)}{e^{2aT}T^{2r}} \rightarrow c(F) = \sum_{\alpha=1}^n \sum_{\beta=1}^n \frac{\varphi_\alpha \bar{\varphi}_\beta l_\alpha l_\beta}{F(\lambda_\alpha) \bar{F}(\lambda_\beta) (i\lambda_\alpha - i\bar{\lambda}_\beta)} > 0,$$

and

$$(3.33) \quad \frac{|\Phi^T(\lambda)F(\lambda)|^2}{(\Phi^T, \Phi^T)} \xrightarrow{e} \frac{1}{(2\pi)^{1/2}c(F)} \left| \sum_{\alpha=1}^n \frac{\varphi_\alpha l_\alpha}{F(\lambda_\alpha)(\lambda_\alpha - \lambda)} \right|^2.$$

For the continuous case,  $a = 0$  and all sequences  $T \rightarrow \infty$

$$(3.34) \quad \frac{(\Phi^T, \Phi^T)}{T^{2r+1}} \rightarrow c(F) = \frac{1}{(2r+1)} \sum_{\alpha=1}^n \left| \frac{\varphi_\alpha}{F(\lambda_\alpha)} \right|^2 > 0,$$

and

$$(3.35) \quad \frac{|\Phi^T(\lambda)F(\lambda)|^2}{(\Phi^T, \Phi^T)} \rightarrow \frac{1}{c(F)(2r+1)} \sum_{\alpha=1}^n \left| \frac{\varphi_\alpha}{F(\lambda_\alpha)} \right|^2 \delta(\lambda - \lambda_\alpha),$$

where  $\delta$  is a delta function.

All integrals involved here can be evaluated by contour integration in the complex plane. Simplifications occur due to the fact that terms contributed by poles at the  $z_\alpha$  and  $\lambda_\alpha$  for  $\alpha = 1, \dots, n$  dominate all others of  $\Phi(z)$  and  $\Phi(\lambda)$  as well as those of  $1/F(z)$  and  $1/F(\lambda)$ .

### THEOREM 3.

(i) For  $a > 0$  and a sequence  $T \rightarrow \infty$  for which  $e^{iT\Omega a} \rightarrow l_a$ , there exists an estimate efficient for spectral densities  $f_1$  and  $f_2$  if and only if

$$(3.36) \quad \left| \frac{F_2(z)}{F_1(z)} \right| = \frac{c(F_1)}{c(F_2)} \left| \frac{\sum_{\alpha=1}^n \frac{\varphi_\alpha l_\alpha z_\alpha}{F_2(z_\alpha)(z_\alpha - z)}}{\sum_{\alpha=1}^n \frac{\varphi_\alpha l_\alpha z_\alpha}{F_1(z_\alpha)(z_\alpha - z)}} \right|$$

for the discrete parameter process, and

$$(3.37) \quad \left| \frac{F_2(\lambda)}{F_1(\lambda)} \right| = \frac{c(F_1)}{c(F_2)} \left| \frac{\sum_{\alpha=1}^n \frac{\varphi_{\alpha} l_{\alpha}}{F_2(\lambda_{\alpha})(\lambda_{\alpha} - \lambda)}}{\sum_{\alpha=1}^n \frac{\varphi_{\alpha} l_{\alpha}}{F_1(\lambda_{\alpha})(\lambda_{\alpha} - \lambda)}} \right|$$

for the continuous parameter process.

(ii) If this condition is satisfied, then any estimate that is efficient for one is also efficient for the other.

PROOF. Under the assumptions of this section  $dF_2(\lambda)/dF_1 = f_2(\lambda)/f_1(\lambda)$  is continuous and  $A = \Lambda$ . For the discrete process  $f_2(z)/f_1(z)$  is always bounded; for the continuous process condition (3.37) implies that  $f_2(\lambda)/f_1(\lambda)$  is bounded above and away from zero. In both cases  $c = c(F_1)/c(F_2) \neq 0$ . Thus Theorems 1 and 2 apply. Expressions (3.36) and (3.37) can be obtained directly from (2.13) by substituting the appropriate forms from Lemma 3.

THEOREM 4.

(i) For  $a = 0$  and any sequence  $T \rightarrow \infty$ , there exists an estimate efficient for  $f_1$  and  $f_2$  if and only if

$$(3.38) \quad f_2(\lambda_{\alpha}) = \frac{c(F_1)}{c(F_2)} f_1(\lambda_{\alpha}) \quad \alpha = 1, \dots, n$$

for both the discrete and the continuous parameter process.

(ii) For the discrete parameter process if (3.38) is satisfied, then any estimate that is efficient for one is efficient for the other.

(iii) For the continuous parameter process if (3.38) is satisfied,  $\bar{k}_0^T$  is an efficient estimate for  $f_1$ , and  $f_2(\lambda)/f_1(\lambda)$  is bounded; then  $\bar{k}_0^T$  is also efficient for  $f_2$ .

PROOF. As before (3.38) is obtained from (2.13) using Lemma 3, and Theorems 1 and 2 apply.

The stronger result of Theorem 3 (ii) is not true in the case  $a = 0$  for the continuous parameter processes, since it is possible to find an efficient estimate for  $f_1$  that depends on derivatives of  $y(t)$  which will not exist for  $f_2$  if the degree of  $1/f_2$  is less than that of  $1/f_1$ . This is, of course, the case when  $f_2/f_1$  is unbounded. However, it is possible to find an estimate that is efficient for all  $f_2$  satisfying (3.38). Such an estimate is given by (3.46).

The case of  $C^T = (0, 1, \dots, T)$  and  $C^T = (0, T)$  can now be treated easily. Under the assumptions made on  $f$  and  $\varphi$ , a solution to the equation (2.2) does exist for both the discrete and the continuous parameter process. (See, for example, Laning and Battin [7], Chapter 8.4.) However, it will not be convenient to use this as the efficient estimates  $n_{ii}^T$  required in the theorems of the previous section. Instead, the "best" estimates for the half-infinite interval will be computed and truncated. The estimates obtained in this way are of some interest and will be given explicitly. For  $a > 0$  and the discrete case let

$$(3.39) \quad \frac{1}{(z - e^{\sqrt{\alpha}})F(z)} = \sum_{s=1}^{\infty} z^{-s} m_{s-1}^{\alpha},$$



then

$$(3.40) \quad \bar{M}^T \bar{k}^T = \sum_{t=0}^T y(T-t) \sum_{\alpha=1}^n \frac{\bar{\varphi}_\alpha e^{-i T \theta \lambda_\alpha} m_1^\alpha}{F(\lambda_\alpha)}.$$

For the continuous case, let

$$(3.41) \quad \frac{1}{(\lambda - \bar{\lambda}_\alpha) F(\lambda)} = E_\alpha(\lambda) + M_\alpha(\lambda)$$

where  $E_\alpha(\lambda)$  is a polynomial

$$(3.42) \quad E_\alpha(\lambda) = \sum_{j=0}^{s-1} e_j^\alpha \lambda^j$$

and  $M_\alpha(\lambda)$  is a proper rational function. Let

$$(3.43) \quad m^\alpha(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda t} M_\alpha(\lambda) d\lambda,$$

then

$$(3.44) \quad \begin{aligned} \bar{M}^T \bar{k}^T = & \sum_{j=0}^{s-1} y^{(j)}(T) \sum_{\alpha=1}^n \frac{\bar{\varphi}_\alpha e^{-i T \theta \lambda_\alpha} (-i)^j e_j^\alpha}{F(\lambda_\alpha)} \\ & + \int_0^T y(T-t) \sum_{\alpha=1}^n \frac{\bar{\varphi}_\alpha e^{-i T \theta \lambda_\alpha} m^\alpha(t)}{F(\lambda_\alpha)} dt. \end{aligned}$$

For  $\alpha = 0$

$$(3.45) \quad \bar{M}^T \bar{k}^T = \sum_{t=0}^T y(t) \sum_{\alpha=1}^n \frac{\bar{\varphi}_\alpha t e^{-i \lambda_\alpha t}}{f(\lambda_\alpha)}$$

and

$$(3.46) \quad \bar{M}^T \bar{k}^T = \int_0^T y(t) \sum_{\alpha=1}^n \frac{\bar{\varphi}_\alpha t e^{-i \lambda_\alpha t}}{f(\lambda_\alpha)} dt.$$

In all cases  $\bar{M}^T$  is a constant to be determined so that the estimate is unbiased; that is,  $\bar{M}^T$  is given by the right side of the expression with  $\varphi(t)$  substituted for  $y(t)$ . A straightforward computation of their variances shows that these estimates are efficient for the half-infinite problem discussed above for all sequences  $T \rightarrow \infty$ . Thus by Lemma 2

$$v_{i0}^T \xrightarrow{c} N_{ii}$$

where  $n_{i0}^T$  indicates the estimates (3.40), (3.44), (3.45), and (3.46) for  $f_i$ , and  $N_{ii}$  are the limit measures given in Lemma 3. The asymptotic forms  $c(F_i)$  also hold for the  $1/f|n_{i0}^T(\lambda)|^2 dF_i(\lambda)$  since

$$(\Phi_i^T, \Phi_i^T) = 1/E_{i0}(T)(n_{i0}^T, n_{i0}^T).$$

Thus Theorems 3 and 4 also hold for  $t = 0, 1, \dots, T$  in the discrete case and  $0 \leq t \leq T$  in the continuous parameter case.

The least square estimate for  $C^T$  half-infinite is given by

$$(3.47) \quad \bar{M}^T \bar{k}_0^T = \sum_{t=-\infty}^T y(t) \bar{\varphi}(t)$$

and

$$(3.48) \quad \bar{M}^T \bar{k}_0^T = \int_{-\infty}^T y(t) \overline{\varphi(t)} dt.$$

For the discrete parameter case,  $F(z) = 1$  provides a bona fide covariance for which  $n_0^T(z)$  for the least square estimate is given by (3.24). Thus from Theorems 3 and 4 in this case the least square estimate is efficient for  $F(z)$  if and only if (3.36) or (3.38) hold for  $F(z) = F_2(z)$  and  $F_1(z) = 1$ . In the continuous case if the least square estimate is efficient then by Lemma 2  $\nu_{ii}^T$  and  $\nu_{i0}^T$  must converge to the same limit.  $N_{ii}$  the limit of  $\nu_{ii}^T$  is given by (3.33) and (3.35) in Lemma 3.  $N_{i0}$  the limit of  $\nu_{i0}^T$  can be computed by use of Lemma 3 and the Helly-Bray Theorem, since  $n_0^T(\lambda)$  is identical with  $\bar{\Phi}^T(\lambda)$  except for a constant where  $\bar{\Phi}^T(\lambda)$  is given by (3.25) with  $F(\lambda) = 1$ . For  $a > 0$  this limit is

$$(3.49) \quad N_{i0}(B) = c \int_B f(\lambda) \left| \sum_{\alpha=1}^n \frac{\varphi_\alpha l_\alpha}{(\lambda_\alpha - \lambda)} \right|^2 d\lambda,$$

and for  $a = 0$  by

$$(3.50) \quad N_{i0}(B) = \frac{1}{\sum_\alpha |\varphi_\alpha|^2 f(\lambda_\alpha)} \int_B \sum_\alpha |\varphi_\alpha|^2 f(\lambda_\alpha) \delta(\lambda - \lambda_\alpha) d\lambda.$$

Thus for  $a = 0$  if the least square estimate is efficient for  $f(\lambda)$  it follows that

$$(3.51) \quad f(\lambda_\alpha) = \text{constant} \quad \alpha = 1, \dots, n.$$

An asymptotic form for  $\int |n_0^T(\lambda)|^2 f(\lambda) d\lambda$  can also be found.

$$(3.52) \quad T^{2r+1} \int |n_0^T(\lambda)|^2 f(\lambda) d\lambda \rightarrow (2r+1) \sum_{\alpha=1}^n |\varphi_\alpha|^2 f(\lambda_\alpha) / \left[ \sum_{\alpha} |\varphi_\alpha|^2 \right]^2.$$

From this and (3.34) of Lemma 3 it is clear that (3.51) is also sufficient.  $N_{i0} = N_{ii}$  for  $a > 0$  becomes

$$(3.53) \quad c' f(\lambda) \left| \sum_{\alpha=0}^n \frac{\varphi_\alpha l_\alpha}{(\lambda_\alpha - \lambda)} \right|^2 = \left| \sum_{\alpha=1}^n \frac{\varphi_\alpha l_\alpha}{F(\lambda_\alpha)(\lambda_\alpha - \lambda)} \right|^2,$$

but this is not possible since  $f(\lambda)$  must be proper. Thus for  $a > 0$  the least square estimate is never efficient.

#### REFERENCES

- [1] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.
- [2] ULF GRENNANDER, "On the estimation of regression coefficients in the case of an auto-correlated disturbance," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 252-272.
- [3] ULF GRENNANDER AND MURRAY ROSENBLATT, "An extension of a theorem of G. Szegő

- and its application to the study of stochastic processes," *Trans. Amer. Math. Soc.*, Vol. 76 (1954), pp. 112-126.
- [4] ULF GRENANDER AND MURRAY ROSENBLATT, "Regression analysis of time series with stationary residuals," *Proc. Nat. Acad. Sci.*, Vol. 40 (1954), pp. 812-816.
  - [5] ULF GRENANDER AND MURRAY ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957.
  - [6] ULF GRENANDER AND GABOR SZEGÖ, *Toeplitz Forms and Their Application*, University of California Press, Berkeley, 1958.
  - [7] J. HALCOMBE LANING, JR. AND RICHARD H. BATTIN, *Random Processes in Automatic Control*, McGraw-Hill Book Co., New York, 1956.
  - [8] MICHEL LOËVE, *Probability Theory*, D. van Nostrand, New York, 1955.
  - [9] HENRY B. MANN AND PAUL B. MORANDA, "On the efficiency of the least square estimates of parameters in the Ornstein Uhlenbeck process," *Sankhyā*, Vol. 13 (1954), pp. 351-358.
  - [10] EMANUEL PARZEN, "Statistical inference on time series by Hilbert space methods, I," Department of Statistics, Stanford University Technical Report No. 23, January 2, 1959.
  - [11] MURRAY ROSENBLATT, "On the estimation of regression coefficients of a vector-valued time series with stationary residuals," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 99-121.
  - [12] MURRAY ROSENBLATT, "Some regression problems in time series analysis," *Proceedings of the Third Berkeley Symposium in Probability and Statistics*, Vol. 1, pp. 165-186, University of California Press, Berkeley, 1956.
  - [13] CHARLOTTE T. STRIEBEL, "On the efficiency of estimates of trend in the Ornstein Uhlenbeck process," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 192-200.
  - [14] BALTHASAR VAN DER POL AND H. BREMMER, *Operational Calculus*, Cambridge University Press, Cambridge, 1955.

## TWO SIMILAR QUEUES IN PARALLEL

BY J. F. C. KINGMAN

*Statistical Laboratory, University of Cambridge*

**1. Introduction.** Haight [3] has considered a system consisting of two unbounded single server queues, in which a customer, on arrival, joins the shorter queue. In the present paper, we make the simplifying assumption of symmetry between the two queues, an assumption that enables us to use generating functions to study the behavior of the stationary solution.

Thus we assume that the two servers each have an exponential service time distribution with unit mean, and that the arrivals form a Poisson process with mean  $2\rho$ . If an arriving customer finds that both queues have equal length, he joins either with probability  $\frac{1}{2}$ .

We first prove that, so long as  $\rho < 1$ , a state of statistical equilibrium is reached. Then the equilibrium equations are converted into an equation for a bivariate generating function, by which this function is given in terms of two univariate generating functions. These two functions are shown to be meromorphic, and the positions of, and residues at, their poles are found. This enables us to express the probabilities as an infinite sum of geometric distributions. It also provides us with approximations valid when  $\rho$  is near unity, such as the result that the waiting time distribution of a customer is the same as that for a single queue with traffic intensity  $\rho^2$ .

**2. Limiting behavior of the system.** The first problem to be decided is whether or not the queue will settle down into a stationary state. Under the assumptions that have been made, the lengths of the two queues form a continuous time Markov process, and we first prove a lemma referring to these processes in general, giving a sufficient condition for a valid limiting distribution to exist. This lemma, which is an extension of a theorem of Foster [2] on the discrete time case, is of wide applicability, and it is hoped to publish an account of further extensions elsewhere.

We consider an irreducible Markov process  $X(t)$ , taking a countable number of values  $i$ , and we assume that the limits

$$q_{ij} = \lim_{t \rightarrow 0} t^{-1} \{P(X(t) = j \mid X(0) = i) - \delta_{ij}\}$$

exist and satisfy the conservation conditions  $\sum_j q_{ij} = 0$ .

LEMMA 1. *Let  $-q_{ii}$  be bounded. Then the limits*

$$p_j = \lim_{t \rightarrow \infty} P(X(t) = j \mid X(0) = i)$$

*exist and are independent of  $i$ . The  $\{p_j\}$  form a probability distribution if and only*

---

Received February 2, 1961; revised June 6, 1961.

if there exist non-negative  $y_i$  such that  $\sum_{j \neq i} q_{ij} y_j < \infty$  for all  $i$ , and

$$(1) \quad \sum_{j \neq i} q_{ij} (y_i - y_j) \geq 1$$

for all but a finite number of  $i$ .

PROOF. Let  $Y(n)$  be a discrete time Markov chain with transition probabilities  $q_{ij}/Q$  ( $i \neq j$ ), where  $Q > -q_{ii}$  for each  $i$ . Then  $Y(n)$  is irreducible, and hence a Césaro limit  $p_j$  of

$$p_{ij}^{(n)} = P(Y(n) = j \mid Y(0) = i)$$

exists as  $n \rightarrow \infty$ , i.e.,

$$p_j = \lim_{n \rightarrow \infty} \frac{1}{n} \left[ \sum_0^n p_{ij}^{(n)} \right].$$

Now define a Poisson process  $N(t)$  with  $N(0) = 0$ , and  $E\{N(t)\} = Qt$ , and let  $X^*(t) = Y\{N(t)\}$ . Then  $X^*(t)$  is a Markov process with the same transition intensities  $q_{ij}$  as  $X(t)$ , and  $X^*(t)$  has, with probability 1, only a finite number of discontinuities in every finite interval. Hence (see, for example, [1]),

$$\begin{aligned} P(X(t) = j \mid X(0) = i) &= P(X^*(t) = j \mid X^*(0) = i) \\ &= P(Y\{N(t)\} = j \mid Y(0) = i) \\ &= \sum_{k=0}^{\infty} \frac{e^{-Qt} (Qt)^k}{k!} p_{ij}^{(k)}. \end{aligned}$$

It follows without difficulty that  $P(X(t) = j \mid X(0) = i) \rightarrow p_j$ , as  $t \rightarrow \infty$ . The  $\{p_j\}$  form a probability distribution if and only if  $Y(n)$  is ergodic. By a result of Foster ([2], Theorem 2) this is so if, and only if, there exist non-negative  $y_i$  such that

$$\sum_{j \neq i} Q^{-1} q_{ij} y_j + (1 + Q^{-1} q_{ii}) y_i < \infty, \quad \text{all } i,$$

and

$$\sum_{j \neq i} Q^{-1} q_{ij} y_j + (1 + Q^{-1} q_{ii}) y_i \leq y_i - 1, \quad \text{all but finitely many } i.$$

This is easily seen to be equivalent to the conditions stated above, and the lemma is proved.

In order to apply this lemma to the problem in hand, we have to consider the values of the  $q_{ij}$  for this process. The  $q_{ij}$  correspond to transitions involving only one event. Thus, if  $m, n$  are the lengths of the two queues, there is a transition of rate  $2\rho$  corresponding to an arrival, which increases the smaller of  $m, n$  by 1. There is also a transition of rate 1 which decreases  $m$  by 1, and another which decreases  $n$  by 1. If we restrict  $y_{mn}$  to be symmetric, we have to satisfy the inequalities

$$2\rho(y_{mn} - y_{m,n+1}) + (y_{mn} - y_{m-1,n}) + (y_{mn} - y_{m,n-1})(1 - \delta_{n0}) \geq 1 \quad (m \geq n)$$

for all but finitely many  $(m, n)$ . It is easily seen that  $y_{mn} = m^2 + n^2$  satisfies these inequalities for sufficiently large  $m$ , so long as  $\rho < 1$ . Hence we obtain

**THEOREM 1.** *There exists a unique limiting distribution  $\{p_{mn}\}$  of the lengths of the two queues so long as  $\rho < 1$ .*

In all the analysis that follows, we shall confine attention to the case  $\rho < 1$ , and to the stationary distribution  $\{p_{mn}\}$ .

**3. The equilibrium equations.** These are derived from the Kolmogorov forward equations in exactly the same way as in Haight's paper [3], and we shall not, therefore, go into the details. We note that, by symmetry,

$$(2) \quad p_{mn} = p_{nm}.$$

With this simplification, the equations become, for all  $m \geq n$ ,

$$(3) \quad \left\{ \begin{array}{l} 2\rho(m = n = 0) \\ 1 + 2\rho(m > n = 0) \\ 2 + 2\rho(m, n \neq 0) \end{array} \right\} p_{mn} = \left\{ \begin{array}{l} 2\rho(m = n) \\ \rho(m = n + 1) \\ 0(m \geq n + 2) \end{array} \right\} p_{m-1,n} + 2\rho p_{m,n-1} + p_{m,n+1} + p_{m+1,n}.$$

Now define

$$(4) \quad F_r(x) = \sum_{n=0}^{\infty} p_{n+r,n} x^n \quad (r \geq 0, |x| \leq 1).$$

Then equations (3) reduce to

$$(5) \quad \begin{cases} x(2\rho x + 1)F_1(x) - (1 + \rho)x F_0(x) = -p_{00}x \\ x(2\rho x + 1)F_2(x) - 2(1 + \rho)x F_1(x) + (1 + \rho x)F_0(x) = p_{00} - p_{10}x \\ x(2\rho x + 1)F_{r+1}(x) - 2(1 + \rho)x F_r(x) + F_{r-1}(x) = p_{r-1,0} - p_{r,0}x \end{cases} \quad (r = 2, 3, \dots)$$

**LEMMA 2.**

$$(6) \quad F(x, y) = \sum_{r=0}^{\infty} F_r(x) y^r$$

exists in  $|x| \leq 1, |y| < 1 + 2\rho$ .

**PROOF.** Put  $x = 1$  in (5) and add the first  $r$  equations.

$$(1 + 2\rho)F_{r+1}(1) - F_r(1) = -p_{r0} \leq 0 \quad (r \geq 1)$$

so that  $F_r(1) \leq (1 + 2\rho)^{1-r} F_1(1)$ . Hence  $|F_r(x)| \leq F_r(1) \leq F_1(1)(1 + 2\rho)^{1-r}$ , and the lemma follows.

In  $|x| \leq 1, |y| < 1 + 2\rho$ , the equations (5) may be combined to give

$$\begin{aligned} x(2\rho x + 1)\{[F(x, y) - F(x, 0)]/y\} - 2(1 + \rho)x F(x, y) + (1 + \rho)x F(x, 0) \\ + y F(x, y) + \rho x y F(x, 0) = (y - x)F(0, y), \end{aligned}$$



or

$$(7) \quad \{x(2\rho x + 1) - 2(1 + \rho)xy + y^2\}F(x, y) \\ = y(y - x)F(0, y) + \{x(2\rho x + 1) - (1 + \rho)xy - \rho xy^2\}F(x, 0).$$

It follows that, whenever  $x$  and  $y$  satisfy  $|x| \leq 1$ ,  $|y| < 1 + 2\rho$ , and

$$(8) \quad x(2\rho x + 1) - 2(1 + \rho)xy + y^2 = 0,$$

then  $y(y - x)F(0, y) = -\{x(2\rho x + 1) - (1 + \rho)xy - \rho xy^2\}F(x, 0)$ , which may be reduced to

$$(9) \quad y(y - x)F(0, y) = -x(2\rho x + 1)\{1 + \rho x - (1 + \rho)y\}F(x, 0).$$

**4. The fundamental correspondence.** We may define a symmetric  $(2 - 2)$  correspondence  $S$  as follows:

DEFINITION.  $Y = Sy$  if, and only if, there exists an  $x$  such that the pairs  $(x, y)$ ,  $(x, Y)$  both satisfy (8). Then  $y + Y = 2(1 + \rho)x$ ,  $yY = x(2\rho x + 1)$ , and, eliminating  $x$ , we obtain

$$(10) \quad \rho Y^2 - 2\{(1 + \rho + \rho^2)y - (1 + \rho)\}Y + y(1 + \rho + \rho y) = 0$$

For a given  $y_0$ , we define an "S-sequence"

$$\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots$$

such that  $y_{n+1} = Sy_n$ ,  $y_{n-1} = Sy_n$ .

LEMMA 3. Any S-sequence  $\{y_n\}$  is of the form

$$(11) \quad y_n = A + \mu(a\lambda^n + a^{-1}\lambda^{-n})$$

where

$$(12) \quad A = (1 + \rho)/2(1 + \rho^2),$$

$\lambda$  is the real, positive root, less than unity, of

$$(13) \quad \lambda + \lambda^{-1} = 2(1 + \rho + \rho^2)/\rho,$$

$$(14) \quad \mu = 2^{-1}\rho^{1/2}/(1 + \rho^2),$$

and  $a$  is an arbitrary complex number.

PROOF. From (10)

$$\rho(y_{n+1} + y_{n-1}) = 2(1 + \rho + \rho^2)y_n - (1 + \rho).$$

Now  $2\rho A = 2(1 + \rho + \rho^2)A - (1 + \rho)$ , so that

$$(y_{n+1} - A) - [(1 + \rho + \rho^2)/\rho](y_n - A) + (y_{n-1} - A) = 0.$$

Hence  $y_n = A + B\lambda^n + C\lambda^{-n}$  for some  $B, C$ . However, since  $y_1 = Sy_0$ , we may put

$$y = A + B + C, \quad Y = A + B\lambda + C\lambda^{-1}$$

in (10), which yields an equation simplifying to

$$BC = \mu^2.$$

Writing  $B = \mu a$ ,  $C = \mu a^{-1}$  proves the lemma.

Two other results which will be used later are:

(i) Since

$$\begin{aligned} \lambda + \lambda^{-1} + 2 &= 2 + 2 \frac{1 + \rho + \rho^2}{\rho} \\ (15) \qquad &= 2(1 + \rho)^2 / \rho = A^2 / \mu^2, \\ \lambda^{\frac{1}{2}} + \lambda^{-\frac{1}{2}} &= A / \mu \end{aligned}$$

(ii) If  $|y| \leq 1 + \rho$ ,  $|Y| \leq 1 + \rho$ , the corresponding value of  $x$  given by  $x = (y + Y) / 2(1 + \rho)$  satisfies

$$|x| \leq \frac{|y| + |Y|}{2(1 + \rho)} \leq 1.$$

**5. The univariate generating function  $F(0, y)$ .** Suppose that  $Y = Sy$ , and that  $|y|, |Y| \leq 1 + \rho$ . Then the corresponding  $x$  has  $|x| \leq 1$ , and we may eliminate  $F(x, 0)$  from (9) to give

$$\begin{aligned} \frac{YF(0, Y)}{yF(0, y)} &= \frac{1 + \rho x - (1 + \rho)Y}{1 + \rho x - (1 + \rho)y} \frac{y - x}{Y - x} = \frac{1 + \rho x - (1 + \rho)Y}{1 + \rho x - (1 + \rho)y} \frac{(1 + 2\rho)x - Y}{(1 + 2\rho)x - y} \\ &= \frac{(1 + \rho x)(1 + 2\rho)x - (1 + \rho)(1 + 2\rho)xY - (1 + \rho x)Y + (1 + \rho)Y^2}{(1 + \rho x)(1 + 2\rho)x - (1 + \rho)(1 + 2\rho)xy - (1 + \rho x)y + (1 + \rho)y^2} \\ &= \frac{\rho x(1 - x) - (1 - x)Y}{\rho x(1 - x) - (1 - x)y} = \frac{Y - \rho x}{y - \rho x} = \frac{(2 + \rho)Y - \rho y}{(2 + \rho)y - \rho Y}. \end{aligned}$$

According to Lemma 3, we may write

$$\begin{aligned} (16) \qquad y &= A + \mu(z + z^{-1}), \\ Y &= A + \mu(\lambda z + \lambda^{-1}z^{-1}), \end{aligned}$$

and we may define

$$(17) \qquad g(z) = yF(0, y).$$

Then

$$\frac{g(\lambda z)}{g(z)} = \frac{(2 + \rho)Y - \rho y}{(2 + \rho)y - \rho Y} = \frac{2A/\mu + \{(2 + \rho)\lambda - \rho\}z + \{(2 + \rho)\lambda^{-1} - \rho\}z^{-1}}{2A/\mu + \{2 + \rho - \rho\lambda\}z + \{2 + \rho - \rho\lambda^{-1}\}z^{-1}}$$

Equation (15) implies that  $z + \lambda^{-1}$  is a factor of both numerator and denominator, so that

$$(18) \qquad \frac{g(\lambda z)}{g(z)} = \frac{\gamma - \lambda^{\frac{1}{2}}z}{\lambda^{\frac{1}{2}}\gamma z - 1},$$

where

$$(19) \quad \gamma = \frac{2 + \rho - \rho\lambda}{\lambda - (2 + \rho)} > 1.$$

Equation (18) is valid in  $|A + \mu(z + z^{-1})|, |A + \mu(\lambda z + \lambda^{-1}z^{-1})| \leq 1 + \rho$ . Now, if  $\lambda^{\frac{1}{2}} \leq |z| \leq \lambda^{-\frac{1}{2}}$ , then

$$|A + \mu(z + z^{-1})| \leq A + \mu|z| + \mu|z|^{-1} \leq A + \mu(\lambda^{\frac{1}{2}} + \lambda^{-\frac{1}{2}}) = 2A = \frac{1 + \rho}{1 + \rho^2} < 1 + \rho.$$

Hence (18) is valid in

$$\lambda^{\frac{1}{2}} - \delta_1 < |\lambda z| < |z| < \lambda^{-\frac{1}{2}} + \delta_2, \quad \text{for some } \delta_1, \delta_2 > 0,$$

and hence in

$$\lambda^{-\frac{1}{2}} - \delta < |z| < \lambda^{-\frac{1}{2}} + \delta, \quad \text{for some } \delta > 0.$$

$g(z)$  is regular in this annulus, and may therefore be expanded in a Laurent series

$$(20) \quad g(z) = \sum_{-\infty}^{\infty} a_n z^n$$

$$\text{Hence } (\lambda^{\frac{1}{2}}\gamma z - 1) \sum a_n \lambda^n z^n = (\gamma - \lambda^{\frac{1}{2}}z) \sum a_n z^n,$$

$$a_n(\lambda^n + \gamma) = a_{n-1}(\lambda^{n-1}\gamma + 1)\lambda^{\frac{1}{2}},$$

$$(21) \quad a_n = a_{-n} = a_0(\lambda^{\frac{1}{2}}\gamma^{-1})^n \prod_{j=1}^n \frac{1 + \lambda^{j-1}\gamma}{1 + \lambda^j\gamma^{-1}}.$$

This defines  $g(z)$  uniquely except for a multiplying factor. Since  $a_n \sim C(\lambda^{\frac{1}{2}}\gamma^{-1})^{|n|}$  as  $|n| \rightarrow \infty$ ,  $g(z)$  is regular in  $\lambda^{\frac{1}{2}}\gamma^{-1} < |z| < \lambda^{-\frac{1}{2}}\gamma$ . Equation (18) may be written as

$$g(\lambda z) = \frac{\gamma - \lambda^{\frac{1}{2}}z}{\lambda^{\frac{1}{2}}\gamma z - 1} g(z),$$

which may be regarded as defining a function regular in  $\lambda^{\frac{1}{2}}\gamma^{-1} < |z| < \lambda^{\frac{1}{2}}\gamma$ , except for a pole at  $z = \lambda^{\frac{1}{2}}\gamma^{-1}$ , and coinciding with  $g(z)$  in  $\lambda^{\frac{1}{2}}\gamma^{-1} < |z| < \lambda^{\frac{1}{2}}\gamma$ . Hence  $g(z)$  can be continued into

$$\lambda^{\frac{1}{2}}\gamma^{-1} < |z| < \lambda^{\frac{1}{2}}\gamma^{-1}$$

except for a pole at  $z = \lambda^{\frac{1}{2}}\gamma^{-1}$ . Repeating this procedure, we can continue  $g(z)$  over the whole unit disc, excluding  $z = 0$ , as a regular function except for poles at

$$z = \lambda^{n+\frac{1}{2}}\gamma^{-1}, \quad (n = 0, 1, 2, \dots).$$

This proves

THEOREM 2.  $F(0, y)$  can be continued to a meromorphic function over the whole

$y$ -plane. Its poles are at the points  $Y_n$ ,  $n = 0, 1, 2, \dots$ , where

$$(22) \quad Y_n = A + \mu(\lambda^{n+1}\gamma^{-1} + \lambda^{-n-1}\gamma)$$

It is an easy matter to show that  $Y_n$  takes its smallest value at  $Y_0 = (2 + \rho)/\rho^2$ . Then, from the fact that, for some  $C$ ,  $F(0, y) - [C/(y - Y_0)]$  is regular in  $|y| < Y_1$ , we obtain the

COROLLARY

$$(23) \quad p_{0n} \sim C'[\rho^2/(2 + \rho)]^n \quad \text{as } n \rightarrow \infty.$$

Let the residue of  $g(z)$  at  $\lambda^{n+1}\gamma^{-1}$  be  $g_n$ . Then  $g(\lambda^{n+1}\gamma^{-1} + \zeta) = g_n/\zeta + O(1)$ . From (18)

$$\frac{g(\lambda^{n+1}\gamma^{-1} + \lambda\zeta)}{g(\lambda^{n+1}\gamma^{-1} + \zeta)} = \frac{\gamma - \lambda^{n+1}\gamma^{-1}}{\lambda^{n+1} - 1} + O(\zeta).$$

Thus

$$\frac{g_{n+1}}{\lambda g_n} = -\gamma \frac{1 - \lambda^{n+1}\gamma^{-2}}{1 - \lambda^{n+1}},$$

so that

$$(24) \quad g_n = g_0(-\lambda\gamma)^n \prod_{j=1}^n \frac{1 - \lambda^j\gamma^{-2}}{1 - \lambda^j}.$$

From this it is easy to see that the residue at  $y = Y_n$  of  $F(0, y)$  is  $\phi_n$ , where

$$(25) \quad \phi_n = \phi_0 \frac{1 - \lambda^{2n+1}\gamma^{-2}}{1 - \lambda\gamma^{-2}} \frac{Y_0}{Y_n} (-\lambda^{-1}\gamma)^n \prod_{j=1}^n \frac{1 - \lambda^j\gamma^{-2}}{1 - \lambda^j}.$$

LEMMA 4. Let  $C_n$  be the contour in the  $y$ -plane corresponding to  $|z| = \lambda^{n+1}$ . Then

$$\sup_{C_n} |y^{-1}F(0, y)| \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

PROOF. Since  $y \sim \mu/z$  as  $z \rightarrow 0$  it is sufficient to prove that

$$G_n = \sup_{\theta} |(\lambda^{n+1}e^{i\theta})^2 g(\lambda^{n+1}e^{i\theta})| \rightarrow 0.$$

From (18),

$$\begin{aligned} \frac{G_n}{G_{n-1}} &\leq \lambda^2 \gamma \sup_{\theta} \left| \frac{1 - \lambda^n \gamma^{-1} e^{i\theta}}{1 - \lambda^n \gamma e^{i\theta}} \right| = \lambda^2 \gamma \frac{1 + \lambda^n \gamma^{-1}}{1 - \lambda^n \gamma}, \\ G_n &\leq G_0 (\lambda^2 \gamma)^n \prod_{j=1}^n \frac{1 + \lambda^j \gamma^{-1}}{1 - \lambda^j \gamma} \rightarrow 0, \quad \text{if } \lambda^2 \gamma < 1. \end{aligned}$$

But, if  $\lambda^2 \gamma \geq 1$ , then

$$\lambda^2(2 + \rho - \rho\lambda) \geq \rho - (2 + \rho)\lambda, \quad (2 + \rho)\lambda(1 + \lambda) \geq \rho(1 + \lambda^2),$$

$$[(2 + \rho)/\rho] \geq \lambda + \lambda^{-1} - 1 = 2[(1 + \rho + \rho^2)/\rho] - 1, \quad 0 \geq \rho.$$

The contradiction establishes the lemma.

THEOREM 3.

$$(26) \quad F(0, y) = F(0, 0) + y \sum_{r=0}^{\infty} \frac{\phi_r}{Y_r(y - Y_r)}.$$

PROOF.  $y^{-1}F(0, y)$  is meromorphic with poles at  $y = 0$  (with residue  $F(0, 0)$ ) and at  $y = Y_r$  (with residue  $\phi_r Y_r^{-1}$ ). By virtue of Lemma 4 we may apply Cauchy's partial fraction theorem to give (26).

COROLLARY

$$(27) \quad p_{n0} = - \sum_{r=0}^{\infty} \phi_r / Y_r^{n+1} \quad (n \geq 1).$$

Putting  $m = n = 0$  in (3) gives

$$(28) \quad p_{00} = \rho^{-1} p_{10}.$$

6. The univariate generating function  $F(x, 0)$ . Equation (9) gives

$$(29) \quad x(2\rho x + 1)F(x, 0) = -y(y - x)F(0, y)/\{1 + \rho x - (1 + \rho)y\}$$

where  $(x, y)$  satisfy (8). Hence  $F(x, 0)$  is an analytic function except when  $y = Y_n$ , or when  $1 + \rho x = (1 + \rho)y$ . This last equation is satisfied only when  $x = 0$  or  $x = 1/\rho^2$ . Now define  $X_n$  as the value of  $x$  such that  $(X_n, Y_{n-1})$  and  $(X_n, Y_n)$  both satisfy (8). Then  $X_0 = 1/\rho^2$ , and it follows that the poles of  $F(x, 0)$  are exactly at

$$(30) \quad x = X_n = (Y_{n-1} + Y_n)/2(1 + \rho) = A(2 + \lambda^n \gamma^{-1} + \lambda^{-n} \gamma)/2(1 + \rho), \quad (n = 0, 1, \dots).$$

Now (29) enables us to find the residue  $\psi_n$  of  $F(x, 0)$  at  $x = X_n$ , namely

$$(31) \quad \psi_n = \frac{A\lambda^{\frac{1}{2}}(X_n - Y_n)(1 - \lambda^{-2n}\gamma^2)g_n}{2(1 + \rho)X_n(2\rho X_n + 1)\{1 + \rho X_n - (1 + \rho)Y_n\}}.$$

It is also clear that the supremum of  $F(x, 0)$  on the contour in the  $x$ -plane corresponding to  $|z| = \lambda^{n+1}$  tends to zero as  $n \rightarrow \infty$ . Hence, as in Theorem 3, we have

THEOREM 4.

$$(32) \quad F(x, 0) = \sum_{r=0}^{\infty} \frac{\psi_r}{x - X_r}$$

$$(33) \quad p_{nn} = - \sum_{r=0}^{\infty} \frac{\psi_r}{X_r^{n+1}}$$

$$(34) \quad \sim C\rho^{2n}$$

as  $n \rightarrow \infty$ , for some  $C$ .

7. The bivariate generating function  $F(x, y)$ . By equation (7),  $F(x, y)$  has singularities only on the planes  $x = X_n$  and  $y = Y_n$ . We may therefore prove

## THEOREM 5.

$$(35) \quad p_{mn} \sim C[\rho^{2m}/(2 + \rho)^{m-n}].$$

as  $m, n \rightarrow \infty$  in  $m \geq n$ , for some  $C$ .

PROOF. Since the nearest singularities to the origin are at  $x = X_0$ , and  $y = Y_0$ ,

$$p_{n+r,n} \sim CX_0^{-n}Y_0^{-r}.$$

Putting  $X_0 = 1/\rho^2$ ,  $Y_0 = (2 + \rho)/\rho^2$  proves the result.

As in the two previous sections, we could make a detailed investigation of the properties of  $F(x, y)$ . However, much of the interest in a queueing system lies with the waiting time distribution, and it will be shown in the next section that this may be determined simply from a knowledge of  $F(x, 0)$ .

**8. The waiting time distribution.** The waiting time of a customer depends on the length of the queue he joins, i.e., on

$$(36) \quad l = \min(m, n).$$

Now  $E(z^l) = \sum_{i=0}^{\infty} z^i \{p_{ii} + 2 \sum_{n=0}^{i-1} p_{in}\} = 2F(z, z) - F(z, 0)$ . In (7) put  $x = y = z$ . Then

$$z(1 - z)F(z, z) = z(1 - z)(1 + \rho z)F(z, 0),$$

so that  $F(z, z) = (1 + \rho z)F(z, 0)$ , and  $E(z^l) = (1 + 2\rho z)F(z, 0)$ . Hence the distribution  $\{p_l\}$  of  $l$  is given by

$$(37) \quad p_l = p_{ll} + 2\rho p_{l-1, l-1},$$

and is determined from Theorem 4.

The distribution of waiting time is then made up of a component of zero waiting time with probability  $p_0$ , together with an absolutely continuous component with density

$$(38) \quad f(W) = \sum_{l=1}^{\infty} p_l \frac{W^{l-1} e^{-W}}{(l-1)!} \quad (W = 0).$$

**9. The one-pole approximation.** It follows from (22) that, for  $n$  large,  $Y_r \sim \mu\gamma\lambda^{-r-1}$ , and from (26) that  $\phi_r \sim C(-\lambda^{-1}\gamma)^r Y_r^{-1}$ . Hence the  $r$ th term in the series (27) for  $p_{n0}$  is of order

$$(\lambda^{n+1}\gamma)^r \quad \text{as } r \rightarrow \infty.$$

For all  $\rho$ ,  $\lambda + \lambda^{-1} \geq 6$ , and hence  $\lambda \leq 3 - 2\sqrt{2} \simeq 0.17$ . Hence, for  $n$  fairly large,  $\lambda^{n+1}\gamma$  will be very small, and we can safely neglect all but a few terms of the series. Even for  $n = 1$  (when  $\lambda^2\gamma$  decreases from 1 to 0.17 as  $\rho$  increases from 0 to 1) this will be valid so long as  $\rho$  is not too small.

Hence, in fairly heavy traffic, we may obtain a reasonable approximation by taking only the first term of the series for  $p_{n0}$ . Similar remarks hold for the other



series, so that

$$(39) \quad \begin{aligned} p_{mn} &\simeq C \frac{\rho^{2m}}{(2 + \rho)^{m-n}} \\ p_{00} &\simeq C \frac{\rho}{2 + \rho} \end{aligned} \quad (m \geq n)$$

for some  $C$ . Equation (37) then shows that

$$(40) \quad p_l \simeq C(1 + 2\rho)\rho^{2l} \quad (l > 0).$$

Thus we are led to

**THEOREM 6.** *In heavy traffic the distribution of waiting time is approximately the same as for a single queue with traffic intensity  $\rho^2$ .*

**10. Related problems.** Haight [3] also considered the case in which a customer is permitted to change queues if by so doing he could improve his position. Under the symmetry conditions that have been imposed in this paper, this process is equivalent, from the point of view of the total number queueing, to a single queue with two servers. The determination of the waiting time distribution is, however, no longer a simple matter, since the order in either queue is not necessarily the order of arrival.

The problem considered in this paper is an example of a random walk on positive integer pairs, with rather complicated boundary conditions. The method of attack used may be generalized to deal with other problems of this sort, and it is hoped to publish an account of this work elsewhere.

This same method, together with the use of the Laplace transform, may also be used to study the transient behavior of the double queue and of other random walks.

**11. Acknowledgments.** I am indebted to Professor D. V. Lindley and Dr. P. Whittle for their helpful suggestions, and to the Department of Scientific and Industrial Research for a research studentship.

#### REFERENCES

- [1] WILLIAM FELLER, "On boundary conditions for the Kolmogorov differential equations," *Ann. Math.*, Vol. 65 (1957), pp. 527-570.
- [2] F. G. FOSTER, "On stochastic matrices associated with various queueing processes," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 355-360.
- [3] F. A. HAIGHT, "Two queues in parallel," *Biometrika*, Vol. 45 (1958), pp. 401-410.

# QUEUES WITH BATCH DEPARTURES I

BY F. G. FOSTER AND K. M. NYUNT

*London School of Economics*

**1. Introduction.** This paper has a pattern closely similar to that of [4]. The following single-server queueing system is considered.

(i) Units arrive at the sequence of instants  $\tau_1, \tau_2, \dots$ , such that the inter-arrival times,  $\theta_n = \tau_{n+1} - \tau_n > 0$  ( $n = 1, 2, \dots$ ), are identically distributed independent random variables with an exponential distribution function,

$$F(x) = P[\theta_n \leq x] = 1 - e^{-\lambda x} \quad (x \geq 0).$$

Put  $\alpha = \int_0^\infty x dF(x)$ . Then  $\lambda = 1/\alpha$ .

(ii) Units are served in batches of exactly  $k$  units by a single server, in order of arrival. Denote by  $\chi_n$  the service time of the  $n$ th batch to be served. We suppose that  $\{\chi_n\}$  ( $n = 1, 2, \dots$ ) is a sequence of identically distributed independent positive random variables, independent also of the sequence  $\{\tau_n\}$ , with common distribution function,  $H(x) = P[\chi_n \leq x]$ . Put  $\psi(s) = \int_0^\infty e^{-sx} dH(x)$ ,  $\beta = \int_0^\infty x dH(x)$  and  $\mu = 1/\beta$ . Define  $\rho = \lambda/\mu$ .

In the terminology of Foster [3], this system can alternatively be described as having the 1-input (arrivals) untriggered with input quantity constantly unity, and an exponential distribution for the 1-input time. The 0-input (departures) is triggered with input quantity constantly,  $k$  and a general distribution for the 0-input time. The system has infinite capacity. For definitions of these concepts, the reader is referred to [3].

Such a batch-size model does not appear to have been treated explicitly in the literature, although it has obvious applications. A simple special case of it is, however, implicit in the work of Jackson and Nichols [5]. These authors suppose that an inter-arrival time devoted to one unit is composed of  $k$  consecutive phases, each exponentially distributed. If instead, we think of this unit as composed of  $k$  subunits (corresponding to the phases of arrival) then we have the idea of batch service: Jackson and Nichols treat the special case of exponential service times.

Justification for the explicit consideration of batch departure systems resides in the fact that the results one can obtain are elegant, and form a natural generalization of the case of unit departures, as treated, for example, in Kendall [6]. The analysis in this paper is similar to that in Bailey [1], but the model is in fact different, and the results obtained here are new. In the terminology of Foster [3], the model Bailey considered differs from the present one in that the 0-input in Bailey's model is untriggered with controlled input quantity of zero to  $k$  units, depending upon the state of the system: the input being, for example, virtual when the system contains no 1's. In other words, service begins from time to time

Received January 9, 1961.

whether or not there happen to be any units in the system. Bailey obtains for this system the equilibrium distribution of queue-size at instants just before service is due to begin.

Denote by  $\xi(t)$ , the number of units in the system, including the batch under service, at the instant  $t$ . Let  $\sigma_n$  be the instant at which the  $n$ th batch (of size  $k$ ) departs from the system on receiving service. Put  $\xi_n = \xi(\sigma_n + 0)$ ,  $n = 1, 2, \dots$ . We shall determine the probability generating function (p.g.f.) of the limiting distribution,

$$p_j^+ = \lim_{n \rightarrow \infty} P[\xi_n = j].$$

The distribution  $\{p_j^+\}$  exists and is independent of the initial state of the system, if, and only if,  $\rho/k < 1$ . The proof of this statement follows the same lines as in the case,  $k = 1$ , as given in Foster [2].

Let us denote this batch departures model by  $E_1/G^k/1$  where  $E_1$  indicates an exponentially distributed inter-arrival time, and  $G^k$  indicates that the service time has a general distribution, and that service is in batches of  $k$  units. We shall consider its relation to the unit departures model,  $E_k/G/1$ , where  $E_k$  indicates an Erlang distribution with parameter  $k$ . We shall derive the equilibrium distribution of queue-size at instants just after departures for this latter system in terms of  $\{p_j^+\}$ . As a special case we shall consider the system,  $E_k/E_r/1$ . In our previous paper [4] we obtained the equilibrium distribution at instants just before arrivals for the same system. In this paper we shall establish the identity of the two formulae, thus verifying a special case of the general proposition that, for the system  $G/G/1$ , when these exist, the equilibrium distribution of queue-size at instants just after departures is identical with that at instants just before arrivals (*cf.* Khintchine [7]).

**2. The system  $E_1/G^k/1$ .** Let  $\{v_n\}$  ( $n = 1, 2, \dots$ ) be a sequence of identically distributed independent random variables with distribution,

$$k_j = P[v_n = j], \quad j = 0, 1, 2, \dots,$$

where

$$k_j = \int_0^\infty \frac{e^{-\lambda x} (\lambda x)^j}{j!} dH(x).$$

Then  $v_n$  is thought of as the number of units joining the queue during the service-time of the  $n$ th batch.

Put  $\kappa(z) = \sum_{j=0}^\infty k_j z^j$ . We note that  $\kappa(z) = \psi\{\lambda(1-z)\}$ . We assume that  $\rho/k < 1$ , and also that  $\kappa(z)$  is regular within the circle  $|z| = 1 + \delta$ , where  $\delta$  is some small positive number. This implies a slight restriction on the distribution,  $H(x)$ , which will always be satisfied in practice. It follows from Rouché's theorem that the equation,

$$(1) \quad \kappa(z) = z^k,$$

has exactly  $k$  roots inside or on the unit circle. For  $\kappa'(1) = \rho < k$ , so that for some small positive  $\delta$ ,  $\kappa(1 + \delta) < (1 + \delta)^k$ . Therefore, on the circle,  $|z| = 1 + \delta$ ,  $|\kappa(z)| = \sum k_j |z|^j < (1 + \delta)^k = |z^k|$ . Clearly,  $z = 1$  is one root, and it is a simple root. Denote the other  $k - 1$  roots by  $\delta_1, \delta_2, \dots, \delta_{k-1}$ .

Define  $P^+(z) = \sum_{j=0}^{\infty} p_j^+ z^j$ .

THEOREM 1.

$$(2) \quad P^+(z) = \frac{(k - \rho)(z - 1) \prod_{j=1}^{k-1} (z - \delta_j) / (1 - \delta_j)}{z^k / \kappa(z) - 1}.$$

PROOF. The process,  $\{\xi_n\}$ , is a Markov process with transition matrix described by the relations:

$$\xi_{n+1} = \max [\xi_n - k, 0] + \nu_n, \quad n = 1, 2, \dots$$

The random variable,  $\max [\xi_n - k, 0]$ , has, in the limit as  $n \rightarrow \infty$ , the generating function,

$$P^+(z)z^{-k} - \sum_{j=0}^{k-1} p_j^+(z^{j-k} - 1).$$

Therefore, since  $\nu_n$  and  $\max [\xi_n - k, 0]$  are independent, we have the relation,

$$P^+(z) = \{P^+(z)z^{-k} - \sum_{j=0}^{k-1} p_j^+(z^{j-k} - 1)\}\kappa(z),$$

which, on simplifying, reduces to

$$P^+(z) = \frac{\sum_{j=0}^{k-1} p_j^+(z^k - z^j)}{z^k / \kappa(z) - 1}.$$

Since  $P^+(z)$  is a probability generating function, it is absolutely convergent in the region,  $|z| \leq 1$ . Therefore, the roots of the numerator in this region must coincide with those of the denominator, and the latter are

$$1, \delta_1, \delta_2, \dots, \delta_{k-1}.$$

Therefore, since the numerator is a polynomial of degree  $k$ , we have

$$\sum_{j=0}^{k-1} p_j^+(z^k - z^j) \equiv C(z - 1) \prod_{j=1}^{k-1} (z - \delta_j),$$

where  $C$  is a constant to be determined. Using  $P^+(1) = 1$ , we find that

$$C = (k - \rho) / \prod_{j=1}^{k-1} (1 - \delta_j)$$

and (2) follows.

If (1) has any roots outside the unit circle, we shall denote them by

$$\beta_1, \beta_2, \dots$$

and we define  $\epsilon_j = 1/\beta_j$ .

EXAMPLE 1. If the service-time distribution is Erlang,  $E_r$ , then

$$\kappa(z) = \{1 + [\rho(1 - z)/r]\}^{-r},$$

and the denominator of (2) becomes

$$z^k \{1 + [\rho(1 - z)/r]\}^r - 1,$$

which, being a polynomial of degree  $k + r$  has precisely the  $(k + r)$  zeros,

$$1, \delta_1, \delta_2, \dots, \delta_{k-1}, \beta_1, \beta_2, \dots, \beta_r,$$

and so can be expressed as

$$C(z - 1) \prod_{j=1}^{k-1} (z - \delta_j) \prod_{j=1}^r (z - \beta_j),$$

where  $C$  is a constant to be determined. Therefore, substituting in (2) and normalizing, we obtain

$$(3) \quad P^+(z) = \prod_{j=1}^r \left( \frac{1 - \beta_j}{z - \beta_j} \right) = \prod_{j=1}^r \left( \frac{1 - \epsilon_j}{1 - \epsilon_j z} \right),$$

where the  $\epsilon_j$ 's are the reciprocals of the roots outside the unit circle of the equation,

$$(4) \quad \{1 + [\rho(1 - z)/r]\}^{-r} = z^k.$$

We can show that these roots, and hence the  $\epsilon_j$ 's are distinct. For suppose, on the contrary, that (4) has a double root, say,  $a$ . Then for  $z = a$ , we should have, by differentiation of (4).

$$(5) \quad \rho \{1 + [\rho(1 - z)/r]\}^{-r-1} = k z^{k-1}.$$

Dividing (4) by (5) and simplifying, we obtain

$$a = (k/\rho)[(r + \rho)/(r + k)].$$

But this value must now satisfy (4); that is

$$(6) \quad \{(r + \rho)/(r + k)\}^{(r+k)/k} = \rho/k.$$

Now we are assuming that the traffic intensity,  $\rho/k$ , is less than unity, say  $\rho/k = 1 - \delta$ , where  $0 < \delta < 1$ . Substituting in (6) and putting  $b = k/(r + k)$ , we get after simplifying,

$$1 - b\delta = (1 - \delta)^b.$$

But this is impossible, unless  $\delta = 0$ . Therefore, (4) has no multiple roots, and so the  $\epsilon_j$ 's are distinct.

It follows that we can write

$$(7) \quad P^+(z) = \sum_{j=1}^r C_j / (1 - \epsilon_j z)$$

where

$$(8) \quad C_j = (1 - \epsilon_j) \prod_{i \neq j} (1 - \epsilon_i) / (1 - \epsilon_i / \epsilon_j).$$

Formula (3) gives the p.g.f. of the equilibrium distribution after departures for the system,  $E_1/E_r^k/1$ . The traffic intensity is  $\rho/k$ . We note here for later use that if the traffic intensity is changed to  $r\rho/k$ , the  $\epsilon_j$ 's become the reciprocals of the distinct roots outside the unit circle of the equation,

$$(9) \quad [1 + \rho(1 - z)]^{-r} = z^k.$$

EXAMPLE 2. If the service-time distribution is exponential,  $E_1$ , we have  $\kappa(z) = \{1 + [\rho(1 - z)]\}^{-1}$ , and (3) becomes

$$(10) \quad P^+(z) = (1 - \epsilon)/(1 - \epsilon z)$$

where  $\epsilon$  is the reciprocal of the root outside the unit circle of

$$(11) \quad \{1 + [\rho(1 - z)]\}^{-1} = z^k.$$

The formula (10) is for the system  $E_1/E_1^k/1$ , with traffic intensity,  $\rho/k$ .

**3. Relationship with the unit departures system  $E_k/G/1$ .** If  $\xi$  is the number of units left by an arbitrary departing batch, then the number of complete batches of size  $k$  in the system at this instant will be

$$\varphi = [\xi/k],$$

where  $[x]$  denotes the greatest integer not greater than  $x$ .

We now interpret the random variable  $\varphi$  as the number of units left by an arbitrary departing unit in the unit departures system  $E_k/G/1$ , which has mean service time  $1/\mu$  and Erlang inter-arrival time distribution,  $E_k$ , with a mean of  $k/\lambda$ . The traffic intensity is thus  $\rho/k$ .

We consider the distribution of  $\varphi$ . Define

$$q_j^+ = P[\varphi = j],$$

and put

$$Q^+(z) = \sum_{j=0}^{\infty} q_j^+ z^j.$$

Now define

$$P_j^+ = \sum_{i=0}^j p_i^+$$

$$Q_j^+ = \sum_{i=0}^j q_i^+.$$

Then

$$(12) \quad \sum_{j=0}^{\infty} P_j^+ z^j = \frac{P^+(z)}{1 - z},$$



and

$$(13) \quad \sum_{j=0}^{\infty} Q_j^+ z^j = \frac{Q^+(z)}{1-z}.$$

We have

$$\begin{aligned} Q_0^+ &= P_{k-1}^+ \\ Q_1^+ &= P_{ik-1}^+ \end{aligned}$$

and generally

$$Q_j^+ = P_{(j+1)k-1}^+ \quad j = 0, 1, \dots.$$

Therefore,

$$\frac{Q^+(z)}{1-z} = \sum_{j=0}^{\infty} P_{(j+1)k-1}^+ z^j.$$

But from (12) we have

$$P_j^+ = \frac{1}{2\pi i} \int_C \frac{P^+(v) dv}{(1-v)v^{j+1}},$$

where  $C$  is a contour around the origin excluding the poles of  $P^+(z)/(1-z)$ . Therefore,

$$\frac{Q^+(z)}{1-z} = \sum_{j=0}^{\infty} \frac{z^j}{2\pi i} \int_C \frac{P^+(v) dv}{(1-v)v^{(j+1)k}},$$

so that

$$(14) \quad Q^+(z) = \frac{1-z}{2\pi i} \int_C \frac{P^+(v) dv}{(1-v)v^k(1-v^{-k}z)}.$$

The poles of the integrand within  $C$  are at

$$v = \omega^j z^{1/k}, \quad j = 1, 2, \dots, k,$$

where  $\omega^j$  is a  $k$ th root of unity. The residue at  $v = \omega^j z^{1/k}$  is

$$1/zk P^+(\omega^j z^{1/k}) \omega^j z^{1/k} / (1 - \omega^j z^{1/k}).$$

Therefore, summing the residues, we obtain the alternative formula,

$$(15) \quad Q^+(z) = \frac{1-z}{zk} \sum_{j=1}^k \frac{P^+(\omega^j z^{1/k}) \omega^j z^{1/k}}{1 - \omega^j z^{1/k}}.$$

EXAMPLE 3. If the service-time distribution is Erlang,  $E_r$ , then  $P^+(z)$  is given by (7), and so from (14),

$$(16) \quad Q^+(z) = \sum_{j=1}^r \frac{1}{2\pi i} \int_C \frac{C_j(1-z) dv}{(1-\epsilon_j v)(1-v)v^k(1-v^{-k}z)}.$$

This is the p.g.f. of the equilibrium distribution of queue-size after departures

in the system  $E_k/E_r/1$ . For traffic intensity  $\rho/k$ , the  $\epsilon_j$ 's are the reciprocals of the roots outside the unit circle of Equation (9).

Now we have

$$\begin{aligned}
 (17) \quad & \frac{1}{2\pi i} \int_C \frac{C_j(1-z) dv}{(1-\epsilon_j v)(1-v)v^k(1-v^{-k}z)} \\
 &= \frac{1}{2\pi i} \int_C \frac{C_j}{(1-\epsilon_j v)(1-v)} \left(1 - \frac{1-v^{-k}}{1-v^{-k}z}\right) dv \\
 &= \frac{1}{2\pi i} \int_C \frac{C_j}{(1-\epsilon_j v)(v-1)} \frac{1-v^{-k}}{1-v^{-k}z} dv,
 \end{aligned}$$

since the neglected part of the integral has no poles inside  $C$ . The integral (17) is most easily evaluated by considering the single pole of the integrand outside  $C$  at  $v = 1/\epsilon_j$ . The residue is

$$-C_j/(1-\epsilon_j)[(1-\epsilon_j^k)/(1-\epsilon_j^k z)].$$

Since the integrand is rational with denominator of degree at least 2 higher than that of the numerator, it follows that (17) is equal to minus the residue outside  $C$ . Therefore, we have

$$(18) \quad Q^+(z) = \sum_{j=1}^r C_j/(1-\epsilon_j)(1-\epsilon_j^k)/(1-\epsilon_j^k z).$$

EXAMPLE 4. If the service-time distribution is exponential,  $E_1$ , we have from (14)

$$\begin{aligned}
 Q^+(z) &= \frac{1-z}{2\pi i} \int_C \frac{1-\epsilon}{1-\epsilon v} \frac{dv}{(1-v)v^k(1-v^{-k}z)} \\
 &= (1-\epsilon^k)/(1-\epsilon^k z)
 \end{aligned}$$

by consideration of the pole at  $v = 1/\epsilon$  outside  $C$ .  $1/\epsilon$  is the single root outside the unit circle of equation (11). This is the p.g.f. of the equilibrium distribution of queue-size after departures in the system  $E_k/E_1/1$  when the traffic intensity is  $\rho/k$ . The formula is, however, found to be identical (apart from notation) with that obtained by Jackson and Nichols [5] for the distribution *before arrivals* in the same system. A more general case of this observation is considered in the next section.

**4. Relationship between the queue-size distributions at departure and at arrival points for the system  $E_k/E_r/1$ .** In our previous paper [4] we obtained the p.g.f.,  $Q(z)$ , for the equilibrium distribution of queue size before arrivals for the system  $E_k/E_r/1$ ,

$$(19) \quad Q(z) = \frac{1-z}{2\pi i} \int_C \frac{P(v) dv}{v(1-v)(1-v^{-r}z)}.$$

$C$  is a contour around the origin excluding the poles of  $P(v)/(1-v)$ , and

$$P(v) = \prod_{j=1}^r \left( \frac{1 - \gamma_j}{1 - \gamma_j v} \right),$$

where for traffic intensity  $r\rho/k$ , the  $\gamma_j$ 's are the roots inside the unit circle of

$$(20) \quad [\rho/(\rho + 1 - z)]^k = z^r.$$

We shall now establish that  $Q(z)$  is identical with  $Q^+(z)$  as given by formula (18) above. We first examine the relationship between the roots,  $1/\epsilon_j$ , of (9) and the roots,  $\gamma_j$ , of (20).

Now if  $1/\epsilon_j$  is a root of Equation (9), then

$$(21) \quad \epsilon_j^k = [1 + \rho(1 - 1/\epsilon_j)]^r.$$

Let us define

$$(22) \quad \gamma_j = 1 + \rho(1 - 1/\epsilon_j).$$

It follows that

$$(23) \quad \epsilon_j = \rho/(\rho + 1 - \gamma_j)$$

and, from (21),

$$(24) \quad \gamma_j^r = \epsilon_j^k.$$

Now from (23) and (24),

$$\gamma_j^r = [\rho/(\rho + 1 - \gamma_j)]^k.$$

But this shows that  $\gamma_j$  is a root of Equation (20), and moreover, from (24),  $|\gamma_j| < 1$ .

The relation (22) thus establishes a one-to-one correspondence between the  $r$  roots,  $1/\epsilon_j$ , of (9) outside the unit circle and the  $r$  roots,  $\gamma_j$ , of (20) inside the unit circle. Since we have proved that the roots,  $1/\epsilon_j$ , are simple, it follows that the poles of the integrand in (19) outside  $C$  are also simple.

From (22), we have

$$\rho(1/\epsilon_i - 1) = 1 - \gamma_i,$$

and

$$\rho(1/\epsilon_i - 1/\epsilon_j) = \gamma_j - \gamma_i.$$

Therefore,

$$(1 - \epsilon_i)/(1 - \epsilon_i/\epsilon_j) = (1 - \gamma_i)/(\gamma_j - \gamma_i).$$

Now define

$$D_j = (1 - \gamma_j) \prod_{i \neq j} (1 - \gamma_i)/(\gamma_j - \gamma_i).$$

It follows that

$$(25) \quad C_j/(1 - \epsilon_j) = \gamma_j^{1-\tau} D_j/(1 - \gamma_j).$$

Analogously to formulae (16) and (17), we can now write

$$(26) \quad Q(z) = \sum_{j=1}^r \frac{1}{2\pi i} \int_C \frac{D_j(1-z) dv}{(1 - \gamma_j v)(1-v)v(1-v^{-\tau}z)}$$

$$(27) \quad = \sum_{j=1}^r \frac{1}{2\pi i} \int_C \frac{D_j}{(1 - \gamma_j v)(v-1)} \frac{v^{\tau-1}(1-v^{-\tau})}{1-v^{-\tau}z} dv.$$

Each integrand in (27) has a single pole outside  $C$  at  $v = 1/\gamma_j$ , and the residue is

$$-\gamma_j^{1-\tau} D_j/(1 - \gamma_j)(1 - \gamma_j^{\tau})/(1 - \gamma_j^{\tau} z)$$

which, by using (24) and (25), we can transform to

$$-C_j/(1 - \epsilon_j)(1 - \epsilon_j^{\tau})/(1 - \epsilon_j^{\tau} z).$$

Since in (27) the sum of the integrands has its denominator of degree higher by 2 than that of the numerator, it follows that  $Q(z) = \sum_{j=1}^r C_j/(1 - \epsilon_j)(1 - \epsilon_j^{\tau})/(1 - \epsilon_j^{\tau} z)$ , which is formula (18) above, and we have proved that, in the system  $E_k/E_r/1$ , the equilibrium distribution of queue size at instants just before arrivals is identical with that at instants just after departures.

**5. Further work.** Let  $\tau_1, \tau_2, \dots$  denote the sequence of instants at which units join the batch departures system. In a sequel we shall consider the existence of the limiting distributions,  $\{p_j^*\}$  and  $\{p_j\}$ , defined, respectively, by

$$p_j^* = \lim_{t \rightarrow \infty} P[\xi(t) = j]$$

and

$$p_j = \lim_{n \rightarrow \infty} P[\xi(\tau_n - 0) = j].$$

We shall also examine the relationships existing between the three distributions

$$\{p_j\}, \{p_j^*\} \text{ and } \{p_j^+\}.$$

#### REFERENCES

- [1] N. T. J. BAILEY, "On queueing processes with bulk service," *J. Roy. Stat. Soc., Ser. B*, Vol. 16 (1954), pp. 80-87.
- [2] F. G. FOSTER, "On the stochastic matrices associated with certain queueing processes," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 355-360.
- [3] F. G. FOSTER, "A unified theory for stock, storage and queue control," *Operat. Res. Quart.*, Vol. 10 (1959), pp. 121-130.
- [4] F. G. FOSTER, "Queues with batch arrivals I," *Acta Math. Acad. Sci. Hung.*, Vol. 12 (1961), pp. 1-10.
- [5] R. R. P. JACKSON AND D. G. NICHOLS, "Some equilibrium results for the queueing process  $E_k/M/1$ ," *J. Roy. Stat. Soc., Ser. B*, Vol. 18 (1956), pp. 275-279.
- [6] D. G. KENDALL, "Some problems in the theory of queues," *J. Roy. Stat. Soc., Ser. B*, Vol. 13 (1951), pp. 151-185.
- [7] A. KHINTCHINE, "Mathematical theory of a stationary queue," *Matem. Sbornik*, Vol. 39 (1932), pp. 73-84. (Russian; German summary).

# NOTES

## ON THE CHAPMAN-KOLMOGOROV EQUATION<sup>1</sup>

BY JACK KARUSH

*University of California, Berkeley*

A partial answer is given to the question of whether every Markov random function comes from a system of transition probabilities satisfying the Chapman-Kolmogorov equation. A given Markov random function determines the transition probabilities up to sets of probability zero and for any choice of the transition probabilities the Chapman-Kolmogorov equation holds up to sets of probability zero. The problem then is one of selecting appropriate versions of the transition probabilities so that the Chapman-Kolmogorov equation holds everywhere. It is shown that such selections exist whenever the time parameter set is countable or whenever the joint distribution of any two of the random variables is absolutely continuous with respect to the product of the marginal distributions. Although the latter condition is always satisfied when the state space is countable, or more generally, when each random variable assumes a countable number of values with probability one, this case, being especially simple, is treated separately. The results are based on exploiting the device of using the marginal distribution when in doubt about what the conditional probability distribution should be.

Let  $(X_t, t \in T)$  be a Markov random function, where  $T$  is a set of real numbers with elements denoted by  $r, s, t, u, v$ . Let  $\mathcal{S}$  be the  $\sigma$ -field of linear Borel sets, and for every  $t$  define  $P_t(S) = P[X_t \in S], S \in \mathcal{S}$ . For every  $s, t, s < t$ , consider the joint probability distribution of  $X_s, X_t$ . There exists what we shall call a version of the conditional probability distribution of  $X_t$  given  $X_s$  or, more concisely, a version of  $P(X_t | X_s)$ , that is, a function  $P_{st}$  of  $x, S, x$  real,  $S \in \mathcal{S}$ , such that  $P_{st}(\cdot, S)$  is Borel for every  $S \in \mathcal{S}$ ,  $P_{st}(x, \cdot)$  is a probability distribution on  $\mathcal{S}$  for every  $x$ , and

$$\int_{\mathcal{S}} P_s(dx) P_{st}(x, S') = P[X_s \in S, X_t \in S'], \quad S, S' \in \mathcal{S}.$$

The Markov property implies that for  $r < s < t$ ,  $P_{rs} * P_{st}$  is a version of  $P(X_t | X_r)$ , where by definition

$$(P_{rs} * P_{st})(x, S) = \int P_{rs}(x, dy) P_{st}(y, S), \quad \text{all } x, S \in \mathcal{S},$$

Received September 26, 1960; revised March 27, 1961.

<sup>1</sup> This paper was prepared with the partial support of the Office of Ordnance Research, U. S. Army under Contract No. DA-04-200-ORD-171, Task Order 3.

so that the Chapman-Kolmogorov ( $C - K$ ) equation

$$P_{rt}(x, S) = (P_{rs} * P_{st})(x, S)$$

holds for  $x \in N \in \mathcal{S}$ ,  $S \in \mathcal{S}$ , where  $P_r(N) = 0$  and  $N$  depends on  $r, s, t, S$ .

On the other hand the usual approach ([1] pp. 89, 255-6) is to start out with  $(P_{st}, s, t \in T, s < t)$  satisfying the  $C - K$  equation identically, together with an arbitrary initial probability distribution  $P_{t_0}$ ,  $T$  being assumed to have a minimum value  $t_0$ , and to construct the probability distribution of the corresponding random functions. A natural question is whether the probability distributions of all Markov random functions with  $T$  having a minimum value are obtained in this manner, or slightly more generally, whether, or under what conditions, one may select versions  $P_{st}$  of  $P(X_t | X_s)$ ,  $s < t$ , satisfying the  $C - K$  equation identically.

Each of the conditions 1-4 below ensures such a selection; 1 and 2 are special cases of 4 and 3, respectively, but are isolated because of their simplicity.

**1.  $T = \text{integers}$ .** In this case an obvious selection is available. For every  $n$  take any version of  $P(X_{n+1} | X_n)$  and define for  $m > 0$ , all  $n$ ,

$$P_{n, n+m}(x, S) = \int P_{n, n+1}(x, dy_1) \int P_{n+1, n+2}(y_1, dy_2) \cdots$$

$$\int P_{n+m-2, n+m-1}(y_{m-2}, dy_{m-1}) P_{n+m-1, n+m}(y_{m-1}, S), \quad \text{all } x, S \in \mathcal{S}.$$

It is easily verified that  $P_{n, n+m}$  is a version of  $P(X_{n+m} | X_n)$  and the  $C - K$  equation is satisfied identically. This amounts to verifying that the operation " $*$ " is associative.

**2. For every  $t$ ,  $P_t$  is discrete, that is, there exists a countable set  $C_t$  such that  $P_t(C_t) = 1$ .** For every  $s$ , if  $P_s(\{x\}) > 0$  then necessarily

$$P_{st}(x, S) = \frac{P[X_s = x, X_t \in S]}{P[X_s = x]}, \quad t > s, S \in \mathcal{S},$$

and if  $P_s(\{x\}) = 0$  define

$$P_{st}(x, S) = P_t(S), \quad t > s, S \in \mathcal{S}.$$

Since  $P_s[x: P_s(\{x\}) = 0] = 0$ ,  $P_{st}$  is a version of  $P(X_t | X_s)$ . If  $r < s < t$  and  $P_r(\{x\}) > 0$  then  $P_{rt}(x, S) = (P_{rs} * P_{st})(x, S)$ ,  $S \in \mathcal{S}$ . If  $P_r(\{x\}) = 0$  then

$$P_{rt}(x, S) = P_t(S) = \int P_s(dy) P_{st}(y, S) = \int P_{rs}(x, dy) P_{st}(y, S) \\ = (P_{rs} * P_{st})(x, S), \quad S \in \mathcal{S}.$$

**3. For every  $s, t, s < t$ , there exists a version  $P_{st}$  such that  $P_{st}(x, \cdot)$  is absolutely continuous with respect to  $P_t(P_{st}(x, \cdot) \ll P_t)$  for all  $x$ , or equivalently, the joint probability distribution of  $X_s, X_t \ll$  the product measure  $P_s \times P_t$ ,**



or equivalently,  $\ll$  some product measure  $\lambda \times \mu$ ,  $\lambda, \mu$   $\sigma$ -finite. We first establish the equivalences. Assume  $P_{st}(x, \cdot) \ll P_t$  for all  $x$  and suppose  $(P_s \times P_t)(B) = 0$ , where  $B$  is a two-dimensional Borel set; then there exists  $N \in \mathcal{S}$  such that  $P_s(N) = 0$  and  $P_t(B_s) = 0$  for  $x \notin N$ , where  $B_s = \{y: (x, y) \in B\}$ , and we have

$$P[(X_s, X_t) \in B] = \int P_s(dx) P_{st}(x, B_s) = 0.$$

Conversely, if the joint probability distribution of  $X_s, X_t \ll \lambda \times \mu$ ,  $\lambda, \mu$   $\sigma$ -finite, then  $P_s \ll \lambda$  and  $P_t \ll \mu$ , so that there exist densities  $dP_s/d\lambda, dP_t/d\mu$ . Let  $S = \{x: (dP_s/d\lambda)(x) > 0\}$ ,  $S' = \{x: (dP_t/d\mu)(x) > 0\}$ . Then  $P_s(S) = 1$ ,  $P_t(S') = 1$ ,  $\lambda \ll P_s$  on  $S$ , and  $\mu \ll P_t$  on  $S'$ , so that  $\lambda \times \mu \ll P_s \times P_t$  on  $S \times S'$  and  $P[(X_s, X_t) \notin S \times S'] = 0$ . It follows that the joint probability distribution of  $X_s, X_t \ll P_s \times P_t$  and therefore has a density which can be taken to be of the form  $p_s(x)p_{st}(x, y)$  where

$$\int p_{st}(x, y) P_t(dy) = 1 \quad \text{for all } x.$$

Then

$$P_{st}(x, S) = \int_S p_{st}(x, y) P_t(dy), \quad \text{all } x, S \in \mathcal{S},$$

defines a version of  $P(X_t | X_s)$  and  $P_{st}(x, \cdot) \ll P_t$  for all  $x$ .

Let  $U$  be the union of a countable dense subset of  $T$  and the countable set of points of  $T$  which are not two-sided limit points of  $T$ . For every  $t$  let

$$N_t = \bigcup_{\substack{t < u < v \\ u, v \in U \\ y \text{ rational}}} [x: P_{tv}(x, S_y) \neq (P_{tu} * P_{uv})(x, S_y)],$$

where  $S_y = (-\infty, y)$ , and define, for  $t < u \in U$ ,

$$\begin{aligned} \hat{P}_{tu}(x, \cdot) &= P_{tu}(x, \cdot) && \text{if } x \notin N_t \\ &= P_u && \text{if } x \in N_t. \end{aligned}$$

Since  $P_t(N_t) = 0$ ,  $\hat{P}_{tu}$  is a version of  $P(X_u | X_t)$  and since a probability distribution on  $\mathcal{S}$  is determined by its values on  $S_y, y$  rational, we have, for  $x \notin N_t$ ,

$$P_{tv}(x, \cdot) = (P_{tu} * P_{uv})(x, \cdot), \quad t < u < v, u, v \in U,$$

and hence

$$\hat{P}_{tv}(x, \cdot) = (\hat{P}_{tu} * P_{uv})(x, \cdot), \quad t < u < v, u, v \in U.$$

For  $x \in N_t$

$$\hat{P}_{tv}(x, \cdot) = (\hat{P}_{tu} * P_{uv})(x, \cdot) = P_v, \quad t < u < v, u, v \in U,$$

by the same reasoning used in 2. Therefore

$$\hat{P}_{tv} = \hat{P}_{tu} * P_{uv}, \quad t < u < v, u, v \in U.$$

Now  $P_{tu}(x, \cdot) \ll P_u$  for every  $x$ ; consequently  $\hat{P}_{tu}(x, \cdot) \ll P_u$  for every  $x$ . It follows that  $\hat{P}_{tu} * P_{us}$  is independent of the version  $P_{us}$  of  $P(X_s | X_u)$  for any  $s > u$ . Let  $P'_{us}$  be another version; then for every  $S \in \mathcal{S}$ ,

$$P_u[y: P_{us}(y, S) \neq P'_{us}(y, S)] = 0$$

so that for every  $x$ ,  $\hat{P}_{tu}(x, [y: P_{us}(y, S) \neq P'_{us}(y, S)]) = 0$  and hence

$$\int \hat{P}_{tu}(x, dy) P_{us}(y, S) = \int \hat{P}_{tu}(x, dy) P'_{us}(y, S)$$

or  $\hat{P}_{tu} * P_{us} = \hat{P}_{tu} * P'_{us}$ . In particular

$$\hat{P}_{tv} = \hat{P}_{tu} * P_{uv} = \hat{P}_{tu} * \hat{P}_{uv}, \quad t < u < v, u, v \in U.$$

If  $s < t \notin U$  there exists a  $u \in U$  such that  $s < u < t$  and we define  $\hat{P}_{st} = \hat{P}_{su} * P_{ut}$ . Then  $\hat{P}_{st}$  is a version of  $P(X_t | X_s)$ , is independent of the version of  $P(X_t | X_u)$  selected, and is well defined, for if  $s < u < v < t$ ,  $u, v \in U$ , we have

$$\hat{P}_{st} * P_{vt} = (\hat{P}_{su} * \hat{P}_{uv}) * P_{vt} = \hat{P}_{su} * (\hat{P}_{uv} * P_{vt}) = \hat{P}_{su} * P_{ut}$$

since  $\hat{P}_{uv} * P_{vt}$  is a version of  $P(X_t | X_u)$ . Finally, the  $\hat{P}_{st}$ 's satisfy the  $C - K$  equation identically. Suppose  $r < s < t$ . If  $s \in U$  then  $\hat{P}_{rt} = \hat{P}_{rs} * \hat{P}_{st}$  by definition. If  $s \notin U$  there exists  $u \in U$  such that  $r < u < s < t$  and, since  $\hat{P}_{us} * \hat{P}_{st}$  is a version of  $P(X_t | X_u)$ ,

$$\hat{P}_{rt} = \hat{P}_{ru} * (\hat{P}_{us} * \hat{P}_{st}) = (\hat{P}_{ru} * \hat{P}_{us}) * \hat{P}_{st} = \hat{P}_{rs} * \hat{P}_{st}.$$

**4.  $T$  is countable.** Here we impose no condition on the  $X_t$ 's, but, guided by 3, we enlarge the exceptional set  $N_s$  to obtain absolute continuity to the extent needed. For every  $s$  define  $N_s$  as above with  $U = T$  and set

$$M_s = N_s \cup \bigcup_{t>s} [x: P_{st}(x, N_t) > 0].$$

Then  $P_s(M_s) = 0$  since  $P_s(N_s) = 0$  and for  $t > s$

$$0 = P_t(N_t) = \int P_s(dx) P_{st}(x, N_t)$$

which implies  $P_s[x: P_{st}(x, N_t) > 0] = 0$ . Suppose  $r < s$  and  $x \notin M_r$ . Then  $P_{rs}(x, M_s) = 0$ ; for  $P_{rs}(x, N_s) = 0$  and if  $t > s$ ,

$$0 = P_{rt}(x, N_t) = \int P_{rs}(x, dy) P_{st}(y, N_t)$$

which implies  $P_{rs}(x, [y: P_{st}(y, N_t) > 0]) = 0$ . For  $s < t$  define

$$\begin{aligned} \hat{P}_{st}(x, \cdot) &= P_{st}(x, \cdot) && \text{if } x \notin M_s \\ &= P_t && \text{if } x \in M_s. \end{aligned}$$

Since  $P_s(M_s) = 0$ ,  $\hat{P}_{st}$  is a version of  $P(X_t | X_s)$ . Suppose  $r < s < t$ . Then arguing separately for  $x \notin M_r$  and  $x \in M_r$ , we obtain for all  $x$ ,  $\hat{P}_{rt}(x, S_y) = (\hat{P}_{rs} * \hat{P}_{st})(x, S_y)$ ,  $y$  rational, so that  $\hat{P}_{rt}(x, \cdot) = (\hat{P}_{rs} * \hat{P}_{st})(x, \cdot)$ . Since

$P_{st}(y, \cdot) = \hat{P}_{st}(y, \cdot)$  if  $y \notin M_s$  and  $\hat{P}_{rs}(x, M_s) = 0$  for all  $x$  it follows that  $\hat{P}_{rt} = \hat{P}_{rs} * \hat{P}_{st}$ .

## REFERENCE

- [1] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.

A GENERALIZATION OF A THEOREM OF BALAKRISHNAN<sup>1</sup>

BY N. DONALD YLVISAKER<sup>2</sup>

*New York University*

**1. Introduction.** Given a stochastic process  $\{X(t), t \in T\}$  on some probability space with second moment kernel

$$\mathcal{E}[X(s)\overline{X(t)}] = K(s, t),$$

a characterization is given of the function

$$m(t) = \mathcal{E}X(t).$$

This characterization includes the result of Balakrishnan [2] for the case of second order stationary, discrete or continuous parameter processes.

**2. The characterization.** Let  $T$  be an abstract set and let  $K$  be a positive definite kernel on  $T \times T$ . A function  $m$  on  $T$  is said to be an admissible mean value function for the kernel  $K$  if there exists a stochastic process  $\{X(t), t \in T\}$  on some probability space with

$$\mathcal{E}[X(s)\overline{X(t)}] = K(s, t) \quad \text{and} \quad \mathcal{E}X(t) = m(t).$$

LEMMA 1.  $m$  is an admissible mean value function for the kernel  $K$  if and only if  $K(s, t) - m(s)\overline{m(t)}$  is positive definite.

PROOF. if  $K(s, t) - m(s)\overline{m(t)}$  is a positive definite kernel on  $T \times T$ , let  $\{X(t), t \in T\}$  be a Gaussian process with mean function  $m$  and covariance kernel  $K(s, t) - m(s)\overline{m(t)}$ , ([3], p. 72). Then

$$\begin{aligned} \mathcal{E}[X(s)\overline{X(t)}] &= \mathcal{E}[X(s) - m(s)][\overline{X(t) - m(t)}] + m(s)\overline{m(t)} \\ &= K(s, t). \end{aligned}$$

Conversely, if  $m$  is admissible,

$$\mathcal{E}[X(s) - m(s)][\overline{X(t) - m(t)}] = K(s, t) - m(s)\overline{m(t)}$$

is positive definite.

Received January 4, 1961; revised May 29, 1961.

<sup>1</sup> This research was sponsored by the Office of Naval Research under Contract Number Nonr 266(33), Project Number NR 042-034, while the author was at Columbia University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

<sup>2</sup> Now at the University of Washington.

To characterize these functions  $m$ , we introduce, for a positive definite kernel  $R$  on  $T \times T$ , the corresponding reproducing kernel Hilbert space of functions on  $T$ , denoted by  $H(R)$ , the dependence on the set  $T$  having been suppressed. For a kernel  $R$ ,  $H(R)$  is specified by the conditions

- (1) for every  $t \in T$ ,  $R(\cdot, t) \in H(R)$ ,
- (2) for every  $t \in T$  and  $f \in H(R)$ ,  $(f, R(\cdot, t))_{H(R)} = f(t)$ .

From these conditions, the following lemma is apparent.

LEMMA 2. Given a function  $m (\neq 0)$  on  $T$ ,  $M(s, t) = m(s)\overline{m(t)}$  is positive definite on  $T \times T$  and  $H(M)$  consists of all multiples of the function  $m$  with  $\|m\|_{H(M)} = 1$ .

We appeal finally to the following general theorem given in [1].

THEOREM 1. Let  $R$  and  $R^*$  be positive definite kernels on  $T \times T$ .  $R - R^*$  is positive definite if and only if  $H(R^*) \subset H(R)$  and for all  $f \in H(R^*)$ ,

$$\|f\|_{H(R^*)} \geq \|f\|_{H(R)}.$$

Returning then to the determination of the functions  $m$  for which  $K(s, t) - m(s)\overline{m(t)}$  is positive definite on  $T \times T$ , we have

THEOREM 2. If  $K$  is a positive definite kernel on  $T \times T$ , then  $K(s, t) - m(s)\overline{m(t)}$  is positive definite if and only if  $m \in H(K)$  and  $\|m\|_{H(K)} \leq 1$ .

That is, the admissible mean value functions for a given second moment kernel  $K$  are those functions in the unit sphere of the reproducing kernel space  $H(K)$ .

Theorem 1 of Balakrishnan may be seen to coincide with Theorem 2 above when  $K$  has the representation

$$K(s, t) = k(s - t) = \int_{-\infty}^{+\infty} \exp[i(s - t)x] dG(x), \quad -\infty < s, t < +\infty.$$

Then, according to Theorem 4D of [4], the unit sphere of  $H(K)$  consists of functions of the form

$$m(t) = \int_{-\infty}^{+\infty} \exp(itx) u(x) dG(x)$$

with

$$\|m\|_{H(K)}^2 = \int_{-\infty}^{+\infty} |u(x)|^2 dG(x) \leq 1.$$

In particular stationary cases, alternative representations are known. Thus, if

$$K(s, t) = \exp[-(s - t)^2/2], \quad -\infty < s, t < +\infty,$$

the unit sphere of  $H(K)$  consists of analytic functions  $m$  for which

$$\sum_{n=0}^{\infty} \frac{1}{n!} \left| \frac{d^n}{dt^n} [\exp(t^2/2)m(t)]_{t=0} \right|^2 \leq 1.$$

It should be noted that Theorem 2 applies even to stationary kernels which do not possess the spectral representation.

Lastly, a nonstationary example is provided by the Brownian motion kernel. For

$$K(s, t) = \min(s, t), \quad 0 \leq s, t \leq 1,$$

the unit sphere of  $H(K)$  consists of absolutely continuous functions  $m$  for which  $m(0) = 0$ , and

$$\int_0^1 |m'(t)|^2 dt \leq 1.$$

- [1] ARONSZAJN, N., "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, Vol. 68 (1950), pp. 337-404.
- [2] BALAKRISHNAN, A. V., "On a characterization of covariances," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 670-675.
- [3] DOOB, J. L., *Stochastic Processes*, New York, John Wiley and Sons, 1953.
- [4] PARZEN, E., "Statistical inference on time series by Hilbert space methods, I," *Tech. Rep. No. 23 (NR-042-993) (1959)*, Appl. Math. and Stat. Lab., Stanford University

## THE OPINION POOL<sup>1</sup>

BY M. STONE<sup>2</sup>

*Princeton University*

**1. Introduction and summary.** When a group of  $k$  individuals is required to make a joint decision, it occasionally happens that there is agreement on a utility function for the problem but that opinions differ on the probabilities of the relevant states of nature. When the latter are indexed by a parameter  $\theta$ , to which probability density functions on some measure  $\mu(\theta)$  may be attributed, suppose the  $k$  opinions are given by probability density functions  $p_{s1}(\theta), \dots, p_{sk}(\theta)$ . Suppose that  $D$  is the set of available decisions  $d$  and that the utility of  $d$ , when the state of nature is  $\theta$ , is  $u(d, \theta)$ .

For a probability density function  $p(\theta)$ , write

$$u[d | p(\theta)] = \int u(d, \theta) p(\theta) d\mu(\theta).$$

The Group Minimax Rule of Savage [1] would have the group select that  $d$  minimising

$$\max_{i=1, \dots, k} \{ \max_{d' \in D} u[d' | p_{si}(\theta)] - u[d | p_{si}(\theta)] \}.$$

As Savage remarks ([1], p. 175), this rule is undemocratic in that it depends only on the *different* distributions for  $\theta$  represented in those put forward by the

Received May 1, 1961; revised August 7, 1961.

<sup>1</sup> Prepared in connection with research sponsored by the Office of Naval Research.

<sup>2</sup> Present address: University College of Wales, Aberystwyth.

group and not on the number of members of the group supporting each different representative.

An alternative rule for choosing  $d$  may be stated as follows: "Choose weights  $\lambda_1, \dots, \lambda_k$  ( $\lambda_i \geq 0$ ,  $i = 1, \dots, k$  and  $\sum_1^k \lambda_i = 1$ ); construct the pooled density function

$$p_{\lambda}(\theta) = \sum_1^k \lambda_i p_{\theta i}(\theta);$$

choose the  $d$ , say  $d_{\lambda}$ , maximising  $u[d | p_{\lambda}(\theta)]$ ." This rule, which may be called the Opinion Pool, can be made democratic by setting  $\lambda_1 = \dots = \lambda_k = 1/k$ .

Where it is reasonable to suppose that there is an actual, operative probability distribution, represented by an 'unknown' density function  $p_a(\theta)$ , it is clear that the group is then acting as if  $p_a(\theta)$  were known to be  $p_{\lambda}(\theta)$ . If  $p_a(\theta)$  were known, it would be possible to calculate  $u[d_{\lambda} | p_a(\theta)]$  and  $u[d_{\theta i} | p_a(\theta)]$ , where  $d_{\theta i}$  is the  $d$  maximising  $u[d | p_{\theta i}(\theta)]$ ,  $i = 1, \dots, k$  and then to use these quantities to assess the effect of adopting the Opinion Pool for any given choice of  $\lambda_1, \dots, \lambda_k$ .

It is of general theoretical interest to examine the conditions under which

$$(1.1) \quad u[d_{\lambda} | p_a(\theta)] \geq \min_{i=1, \dots, k} u[d_{\theta i} | p_a(\theta)].$$

Theorems 2.1 and 3.1 provide different sets of sufficient conditions for (1.1) to hold. Theorem 2.1 requires  $k = 2$  and places a restriction on  $p_a(\theta)$  (or, equivalently, on  $p_{\theta 1}(\theta)$  and  $p_{\theta 2}(\theta)$ ); Theorem 3.1 puts conditions on  $D$  and  $u(d, \theta)$  instead.

**2. The case of  $k = 2$ .** The following example shows that conditions are needed for (1.1) to hold. With  $k = 2$ , suppose that  $p_{\theta 1}(\theta)$ ,  $p_{\theta 2}(\theta)$ ,  $p_a(\theta)$  are given by atoms of probability one on  $\theta_1$ ,  $\theta_2$ ,  $\theta_a$  respectively, where  $\theta_1$ ,  $\theta_2$ ,  $\theta_a$  are different; also suppose that  $D$  has only three elements  $d_1$ ,  $d_2$ ,  $d_3$  and that

$$\begin{aligned} u(d_1, \theta_1) &= 1, & u(d_2, \theta_1) &= 0, & u(d_3, \theta_1) &= \frac{3}{4}, \\ u(d_1, \theta_2) &= 0, & u(d_2, \theta_2) &= 1, & u(d_3, \theta_2) &= \frac{3}{4}, \\ u(d_1, \theta_a) &= \frac{1}{2}, & u(d_2, \theta_a) &= \frac{1}{2}, & u(d_3, \theta_a) &= 0. \end{aligned}$$

Then  $d_{\lambda 1} = d_1$ ,  $d_{\lambda 2} = d_2$  and, for  $\lambda_1 = \lambda_2 = \frac{1}{2}$ ,  $d_{\lambda} = d_3$  and (1.1) does not obtain.

However, the following theorem may be stated:

**THEOREM 2.1.** *If, for some  $\mu_1, \mu_2$ ,  $p_a(\theta) = \mu_1 p_{\theta 1}(\theta) + \mu_2 p_{\theta 2}(\theta)$ , then (1.1) holds for any weights  $\lambda_1, \lambda_2$ . (As heretofore explicit, the assumption is made that  $d_{\lambda 1}$ ,  $d_{\lambda 2}$ ,  $d_{\lambda}$  exist.)*

**PROOF.**  $d_{\theta i}$  maximises  $u[d | p_{\theta i}(\theta)]$ ,  $i = 1, 2$ , and  $d_{\lambda}$  maximises  $u[d | p_{\lambda}(\theta)]$  or  $\lambda_1 u[d | p_{\theta 1}(\theta)] + \lambda_2 u[d | p_{\theta 2}(\theta)]$ . Writing  $b_{ij}$  for  $u[d_{\theta i} | p_{\theta j}(\theta)] - u[d_{\lambda} | p_{\theta j}(\theta)]$ , it follows that



- (2.1)  $b_{11} \geq 0,$   
 (2.2)  $b_{22} \geq 0,$   
 (2.3)  $\lambda_1 b_{11} + \lambda_2 b_{12} \leq 0,$   
 (2.4)  $\lambda_1 b_{21} + \lambda_2 b_{22} \leq 0.$

For (1.1) to hold, it is necessary that either

- (2.5)  $\mu_1 b_{11} + \mu_2 b_{12} \leq 0$  or  
 (2.6)  $\mu_1 b_{21} + \mu_2 b_{22} \leq 0.$

Now it is necessary that  $\mu_1 + \mu_2 = 1$  so that, if  $\mu_1 \leq \lambda_1$ , (2.1) and (2.3) imply (2.5); while, if  $\mu_1 > \lambda_1$ , (2.2) and (2.4) imply (2.6). Therefore (1.1) holds and the theorem is established.

EXAMPLE. If each of  $p_a(\theta)$ ,  $p_{s1}(\theta)$ ,  $p_{s2}(\theta)$  is atomic on two  $\theta$ -points and if  $p_{s1}(\theta)$ ,  $p_{s2}(\theta)$  are not identical,  $p_a(\theta)$  may be written as  $\mu_1 p_{s1}(\theta) + \mu_2 p_{s2}(\theta)$  and (1.1) obtains. If  $p_{s1}(\theta) = p_{s2}(\theta)$ , (1.1) clearly obtains.

**3. The general case.** That the condition  $p_a(\theta) = \mu_1 p_{s1}(\theta) + \dots + \mu_k p_{sk}(\theta)$  is not sufficient for (1.1), when  $k > 2$ , follows from the following example: Suppose that  $k = 3$  and that  $p_{si}(\theta)$  is given by an atom of probability one at  $\theta = \theta_i$  for  $i = 1, 2, 3$  where  $\theta_1, \theta_2, \theta_3$  are different; also suppose that  $D$  has only four elements  $d_0, d_1, d_2, d_3$  for which

$$\begin{aligned} u(d_0, \theta_1) &= \frac{3}{4}, & u(d_1, \theta_1) &= 2\frac{1}{2}, & u(d_2, \theta_1) &= \frac{1}{4}, & u(d_3, \theta_1) &= \frac{1}{4}, \\ u(d_0, \theta_2) &= \frac{3}{4}, & u(d_1, \theta_2) &= \frac{1}{4}, & u(d_2, \theta_2) &= 2\frac{1}{2}, & u(d_3, \theta_2) &= \frac{1}{4}, \\ u(d_0, \theta_3) &= 0, & u(d_1, \theta_3) &= \frac{1}{4}, & u(d_2, \theta_3) &= \frac{1}{4}, & u(d_3, \theta_3) &= 2\frac{1}{2}. \end{aligned}$$

Choose a small positive number  $\epsilon$ . Suppose  $[\mu_1, \mu_2, \mu_3]$  is such that  $p_a(\theta)$  is atomic on  $[\theta_1, \theta_2, \theta_3]$  with

$$[p_a(\theta_1), p_a(\theta_2), p_a(\theta_3)] = [\frac{1}{3}(1 - \frac{1}{2}\epsilon), \frac{1}{3}(1 - \frac{1}{2}\epsilon), \frac{1}{3}(1 + \epsilon)].$$

Take  $[\lambda_1, \lambda_2, \lambda_3]$  so that  $p_{s1}(\theta)$  is atomic on  $[\theta_1, \theta_2, \theta_3]$  with

$$[p_{s1}(\theta_1), p_{s1}(\theta_2), p_{s1}(\theta_3)] = [\frac{1}{3}(1 + \frac{1}{2}\epsilon), \frac{1}{3}(1 + \frac{1}{2}\epsilon), \frac{1}{3}(1 - \epsilon)].$$

Then  $u[d_0 | p_{s1}(\theta)] = 1 + \frac{1}{2}\epsilon$ ,  $u[d_1 | p_{s1}(\theta)] = u[d_2 | p_{s1}(\theta)] = 1 + 9\epsilon/24$ ,  $u[d_3 | p_{s1}(\theta)] = 1 - 3\epsilon/4$ ; whence  $d_{s1} = d_0$ . Also, by symmetry,  $u[d_0 | p_a(\theta)] = 1 - \frac{1}{2}\epsilon$ ,  $u[d_1 | p_a(\theta)] = u[d_2 | p_a(\theta)] = 1 - 9\epsilon/24$ ,  $u[d_3 | p_a(\theta)] = 1 + 3\epsilon/4$ ; whence

$$u[d_{s1} | p_a(\theta)] = u[d_0 | p_a(\theta)] < \min \{u[d_{si} | p_a(\theta)] | i = 1, 2, 3\}$$

so that (1.1) does not hold.

Theorem 2.1 gives conditions on  $k$  and  $p_a(\theta)$  for (1.1) to obtain. The following theorem gives conditions on only  $D$  and  $u(d, \theta)$  for (1.1) to obtain.

**THEOREM 3.1.** If (i)  $D$  is an interval of real numbers (ii)  $-u(d, \theta)$  is, for each  $\theta$ , a strictly convex function of  $d$  then (1.1) holds for all weights  $\lambda_1, \dots, \lambda_k$ . (The assumption is made that  $d_{s1}, \dots, d_{sk}, d_{s\lambda}$  exist.)

**PROOF.** Consider any three different elements  $d_1, d_2, d_3$  of  $D$  such that  $d_1 = \rho d_2 + (1 - \rho)d_3$ ,  $0 < \rho < 1$ . Then, for all  $\theta$ ,  $u(d_1, \theta) > \rho u(d_2, \theta) + (1 - \rho)u(d_3, \theta)$  and hence  $u[d_1 | p(\theta)] > \rho u[d_2 | p(\theta)] + (1 - \rho)u[d_3 | p(\theta)]$ . Therefore  $-u[d | p_a(\theta)]$ ,  $-u[d | p_{si}(\theta)]$ ,  $i = 1, \dots, k$ , are strictly convex in  $d$ . Let  $d_m = \min \{d_{s1}, \dots, d_{sk}\}$  and  $d_M = \max \{d_{s1}, \dots, d_{sk}\}$ . For  $d_m \leq d \leq d_M$ , by the convexity of  $-u[d | p_a(\theta)]$ ,

$$(3.1) \quad u[d | p_a(\theta)] \geq \min \{u[d_m | p_a(\theta)], u[d_M | p_a(\theta)]\}.$$

Hence

$$(3.2) \quad \min_{i=1, \dots, k} u[d_{si} | p_a(\theta)] = \min \{u[d_m | p_a(\theta)], u[d_M | p_a(\theta)]\}.$$

For weights  $\lambda_1, \dots, \lambda_k$ , if  $d_m \leq d_{s\lambda} \leq d_M$ , (3.1) and (3.2) together imply (1.1). However, if  $d_{s\lambda} < d_m$ , there exists a  $d^* \in D$  and  $\rho_i^*$ ,  $0 < \rho_i^* < 1$ ,  $i = 1, \dots, k$ , such that  $d_{s\lambda} < d^* < d_m$  and  $d^* = \rho_i^* d_{s\lambda} + (1 - \rho_i^*) d_{si}$ ,  $i = 1, \dots, k$ . By the established strict convexities,

$$\begin{aligned} u[d^* | p_{si}(\theta)] &> \rho_i^* u[d_{s\lambda} | p_{si}(\theta)] + (1 - \rho_i^*) u[d_{si} | p_{si}(\theta)] \\ &\geq \rho_i^* u[d_{s\lambda} | p_{si}(\theta)] + (1 - \rho_i^*) u[d_{s\lambda} | p_{si}(\theta)] \\ &= u[d_{s\lambda} | p_{si}(\theta)], \end{aligned} \quad i = 1, \dots, k;$$

whence  $\sum_i \lambda_i u[d^* | p_{si}(\theta)] > \sum_i \lambda_i u[d_{s\lambda} | p_{si}(\theta)]$  or

$$u[d^* | p_{s\lambda}(\theta)] > u[d_{s\lambda} | p_{s\lambda}(\theta)],$$

a contradiction. Hence  $d_{s\lambda} < d_m$  is impossible; and so is  $d_M < d_{s\lambda}$ . Therefore the theorem is established.

**EXAMPLE.**  $D$  is an interval,  $\theta$  is a real parameter and  $u(d, \theta) = -(d - \theta)^2$ . Because  $(d - \theta)^2$  is strictly convex in  $d$  for each  $\theta$ , (1.1) obtains.

In conclusion, it may be noted that it is quite possible to have

$$u[d_{s\lambda} | p_a(\theta)] > \max \{u[d_{si} | p_a(\theta)] | i = 1, \dots, k\}.$$

For example, this will occur (for all but degenerate cases) when

$$p_a(\theta) = \sum_i \mu_i p_{si}(\theta)$$

and  $\lambda_i = \mu_i$ ,  $i = 1, \dots, k$ .

#### REFERENCE

- [1] L. J. SAVAGE, *The Foundations of Statistics*, John Wiley and Sons, New York, 1954.

# CORRECTION NOTES AND ACKNOWLEDGMENTS OF PRIORITY

## CORRECTIONS TO "STATISTICAL METHODS IN MARKOV CHAINS"

BY PATRICK BILLINGSLEY

*The University of Chicago*

E. S. Keeping has pointed out a mistake in the proof of Theorem 2.1 of the above-titled paper (*Ann. Math. Stat.*, Vol. 32 (1961), pp. 12-40). The error can be corrected by making the following changes. Replace the display preceding (2.4) by

$$N_{uv}^{(n)}(F) = \sum_v N_{uv}^{(n-1)}(F(w, v)),$$

and in the following line replace  $f_{uv} > 0$  by  $f_{vv} > 0$ . Replace (2.4) by

$$F_{vv}^* = \sum_w f_{vw} f_w^{-1} F_{wu}^*(w, v).$$

In line 1, page 15, replace  $F^*(u, w)$  by  $F^*(w, v)$ , "column" by "row", and

$$F_{vw}^*(u, w) = F_{vw}^*$$

by  $F_{vu}^*(w, v) = F_{vu}^*$ . In line 3, replace  $\sum_w f_{uw} F_{vw}^*$  by  $\sum_w f_w^* F_{vu}^*$  in two places.

*Misprints:* p. 13, line 27, for 1 read  $p_i$ ; p. 22, line 19, for  $i$  read 1; a factor of 2 is missing on the right in the first display on p. 26, and on the left in (5.4), (5.5), and (5.6).

---

## CORRECTION TO

## "A CONSERVATIVE PROPERTY OF BINOMIAL TESTS"

BY H. A. DAVID

*Virginia Polytechnic Institute*

In the proof of inequality (1) of the above note (*Ann. Math. Stat.*, Vol. 31 (1960), pp. 1205-1207) it is tacitly assumed, near equation (4), that  $P$  can be a maximum only if  $\Pr(S_{n-1} \geq a_n)$  is a maximum for any given  $\pi_n$ . I am indebted to Dr. W. Hoeffding for pointing this out. His proof, cited in my note, establishes the inequality without such an assumption.

---

## ACKNOWLEDGMENT OF PRIORITY

BY V. P. GODAMBE

*Science College, Nagpur, India*

In connection with my article "An optimum property of maximum likelihood estimation" (*Ann. Math. Stat.*, Vol. 31 (1960), pp. 1208-1211), I wish to ac-

knowledge that Professor S. S. Wilks proved a special case of my theorem in his article "Shortest average confidence intervals from large samples," (*Ann. Math. Stat.*, Vol. 9 (1938), p. 172), which was overlooked during my research.

---

### CORRECTION TO

#### "A PROOF OF WALD'S THEOREM ON CUMULATIVE SUMS"

BY N. L. JOHNSON

*University College, London*

The following correction should be made in the above-titled article (*Ann. Math. Stat.*, Vol. 30 (1959), pp. 1245-7): On p. 1245,  $\varepsilon(n) < \infty$  should be added to condition (iii) of Theorem 1.

---

### CORRECTION TO

#### "ON THE MUTUAL INDEPENDENCE OF CERTAIN STATISTICS"

BY C. G. KHATRI

*University of Baroda*

I am indebted to Robert Wijsman for calling my attention to the following misprints in the above mentioned paper (*Ann. Math. Stat.*, Vol. 30 (1959), pp. 1258-1262).

(i) Page 1258, the last two lines of (2.2) should read "the elements below the principal diagonals are

$$t_{ji} = |A_{ji}| / (|A_{ii}| \cdot |A_{i-1, i-1}|)^{\frac{1}{2}},$$

$$A_{ji} = \begin{pmatrix} s_{11} & \cdots & s_{1i} \\ \cdots & \cdots & \cdots \\ s_{i-1,1} & \cdots & s_{i-1,i} \\ s_{j1} & \cdots & s_{ji} \end{pmatrix} : i \times i (j \geq i)$$

and the vertical bars on both sides of a matrix denote the determinant of that matrix."

(ii) Page 1259, the reference of (2.7) should be "Ingram Olkin, Institute of Statistics Mimeographed Series No. 43 (1951), p. 74, Corollary 1.30" instead of [3].

(iii) Page 1259, replace the line after (2.8) by "where  $dX = \prod_{ij} dx_{ij}$  and  $dS = \prod_{i \geq j} ds_{ij}$ ."

(iv) Page 1259, Section 3, line 4, replace the last  $S$  by  $S_1$ .

(v) Page 1259, Section 3, line 5, the matrix  $I - ZZ'$  should be in vertical bars rather than in parentheses.

**CORRECTION AND ACKNOWLEDGMENT OF PRIORITY TO  
"FIRST PASSAGE TIMES FOR A GENERALIZED RANDOM WALK"**

BY JOHN R. KINNEY

*Massachusetts Institute of Technology*

The following correction should be made in the above-titled article (*Ann. Math. Stat.*, Vol. 32 (1961), pp. 235-243). The equations on line 13, p. 239 should read

$$L^{-1}M = L^{-1}MH + K \quad \text{so} \quad G^* = G^*H + K.$$

At the time the above paper was published the author was unaware that Theorem 2 was a version of Wald's Identity as shown in the following reference (A. Wald, "On cumulative sums of random variables," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 283-296).

---

**CORRECTION TO**

**"TABLES OF EXPECTED VALUES OF ORDER STATISTICS AND  
PRODUCTS OF ORDER STATISTICS FOR SAMPLES OF SIZE  
TWENTY AND LESS FROM THE NORMAL DISTRIBUTION"**

EDITED BY D. TEICHROEW

*Stanford University*

The following correction should be made in the above-titled paper (*Ann. Math. Stat.*, Vol. 27 (1956), pp. 410-426). On p. 416, for  $N = 10$ ,  $i = 3$ , the last digit of the expectation should be 4, not 7; thus the expression should read .65605 91054.

## ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Central Regional Meeting of the Institute, Urbana, Illinois, November 24-25, 1961. Additional abstracts will appear in the March, 1962 issue.)

### 1. Admissibility of the Optimal Invariant Estimate for a Translation Parameter Under Absolute Error Loss Function. MARTIN FOX AND HERMAN RUBIN, Michigan State University.

Let  $P$  satisfy the conditions given by Stein (*Ann. Math. Stat.*, Vol. 30, pp. 970-979) with the following changes: (i) replace Stein's condition (2.6) with the condition

$$\int d\nu(y) \int x^2 d_x P(x, y) < \infty$$

and (ii) add the condition that the unique median of  $P(\cdot, y)$  be at 0 for each  $y \in \mathcal{Y}$ . Then, with absolute error loss function,  $x$  is an almost admissible estimate of the translation parameter  $\theta$  where  $(X - \theta, Y)$  has the joint distribution  $P$ . The proof is similar to Stein's but somewhat more intricate. Furthermore, under the assumption that  $p(\cdot, y)$  is a density for each  $y \in \mathcal{Y}$ , Stein's proof of admissibility goes through. An example shows that almost admissibility is the best that can be obtained without a density. Farrell (Ithaca meeting, April 20-22, 1961) has shown that, if the median is nonunique, then there is no admissible estimate. The results stated above are still valid if the loss function is weighted by  $a$  if the estimate is to the left of  $\theta$  and by  $b$  otherwise. The condition on the median is replaced by the same condition for the  $(1 - \alpha)$ th quantile where  $\alpha = a/(a + b)$ .

### 2. Unbiased Estimation of Probability Densities (Preliminary report). S. G. GHURYE, University of Minnesota.

Let  $y = (y_1, \dots, y_n)$  be a sample from a  $k$ -dimensional population  $P$ , which is an element of a family,  $\mathcal{P}$ , of distributions. Let  $g(x)$  be a known numerical-valued function on  $R_k$  with finite expectation  $\omega_P = \int_{R_k} g dP$ , for all  $P \in \mathcal{P}$ . It is desired to find an unbiased estimate  $\Phi(y)$  of  $\omega_P$ . If  $\mathcal{P}$  is a dominated family with respective probability densities  $f_P(x)$  relative to a known measure  $\mu$  on the  $k$ -dimensional Borel sets, then the problem is equivalent to that of finding  $\varphi(x, y)$  satisfying  $E_P^y \varphi(x, y) = f_P(x)$  for all  $x \in R_k$ ,  $P \in \mathcal{P}$ . A number of special cases of the problem have been treated previously [Kolmogorov, *Izvest. Akad. Nauk SSSR, Ser. Mat.* (1950); Lieberman and Resnikoff, *J. Amer. Stat. Assoc.* (1955); Washio, Morimoto and Ikeda, *Bull. Math. Stat.* (1956); Schmetterer, *Ann. Math. Stat.* (1960)]. We give a detailed discussion for many families of densities, and also consider certain functions of densities.

### 3. On the Resolution of Statistical Hypotheses. ROBERT V. HOGG, University of Iowa.

Let  $\omega_0$  be the space of a parameter  $\theta$ . Let  $\omega_i$  be a subset of  $\omega_{i-1}$ ,  $i = 1, 2, \dots, k$ . We test  $\theta \in \omega_k$  against  $\theta \in \omega_0 - \omega_k$  by testing iteratively the following hypotheses:  $\theta \in \omega_i$  against  $\theta \in \omega_{i-1} - \omega_i$ ,  $i = 1, 2, \dots, k$ . The hypothesis  $\theta \in \omega_k$  is accepted if and only if each intermediate hypothesis is accepted. If the test statistic for each intermediate hypothesis  $\theta \in \omega_i$  is based on the corresponding likelihood ratio  $\lambda_i$ , we demonstrate why, under fairly general conditions, these test statistics are mutually stochastically independent. This argument is based on an independence theorem which deals with complete sufficient statistics.



4. **A  $3(2^{8-4})$  Design of Resolution V.** PETER W. M. JOHN, University of California, Davis. (By title)

The smallest  $2^{k-p}$  fraction of the  $2^8$  design of resolution V (main effects and two factor interactions clear) is the quarter replicate, involving 64 points. A three-sixteenth replicate, 48 points, of resolution V is obtained, in which each of the main effects and two factor interactions is estimable from at least a combination of two of the sixteenths. In the three-quarter replicate of the  $2^8$  design,  $3(2^{8-4})$ , obtained by omitting the quarter replicated defined by  $I = ABC = DEF = ABCDEF$ , put  $G = ABDE$  and  $H = ACDF$  to form the  $3(2^{8-4})$  design. The three sixteenths may be combined in pairs to give the following eighth replicates from which the desired estimates are obtained. They are the fractions whose defining contrast subgroups are generated by (i)  $ABC, ABDEG, ACDFH$ ; (ii)  $DEF, ABDEG, ACDFH$ ; (iii)  $ABCDEF, ABDEG, ACDFH$ .

5. **The Bivariate Chi Distribution.** P. R. KRISHNAIAH, PETER HAGIS, JR. AND LEON STEINBERG, Remington Rand Univac, Blue Bell, Penna.

Consider a vector  $\mathbf{X}_j = (X_{1j}, X_{2j})$  of two random variables whose joint distribution is the bivariate normal with mean vector  $\boldsymbol{\mu} = (\mu_1, \mu_2)$  and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

If we let  $\chi_1 = [\sum_{j=1}^n (X_{1j}^2/\sigma_1^2)]^{1/2}$  and  $\chi_2 = [\sum_{j=1}^n (X_{2j}^2/\sigma_2^2)]^{1/2}$ , then the joint distribution of  $\chi_1$  and  $\chi_2$  is called the central or non-central bivariate chi distribution according to whether  $\boldsymbol{\mu} = \mathbf{0}$  or  $\boldsymbol{\mu} \neq \mathbf{0}$ . In the present paper, some properties of this distribution are discussed. Also, a test is proposed to test for the homogeneity of mean lives of machines when the failure times follow a bivariate chi or chi-square distribution with known correlation. The monotonicity property of the power of this test is established and extensive tables are constructed for use in applications of the test. Applications of the test in areas other than life testing are also discussed.

6. **On the Efficiency of Optimum Nonparametric Procedures in the Two Sample Case.** P. W. MIKULSKI, University of California, Berkeley.

Consider the hypothesis that two samples are drawn from two populations with the same, continuous, completely specified distribution  $F$ . The alternative is that the distribution in the second population is shifted to the right. For testing this hypothesis consider: (a) the locally most powerful rank test, and (b) the locally asymptotically optimal test (J. Neyman: "Optimal Asymptotic Tests of Composite Statistical Hypotheses, *H. Cramer Volume*, 1959), or the asymptotically equivalent likelihood ratio test. The question is investigated, how the Pitman efficiency  $e(F; \Psi)$  of the optimal rank test to the optimal parametric test behaves, if the true distribution  $\Psi$  departs from the assumed distribution  $F$ . Chernoff and Savage have shown that if  $F$  is normal, then  $e(F; \Psi) \geq 1$  for all  $\Psi$ . It turns out that under certain regularity restrictions for the assumed distribution, normality of  $F$  is a necessary condition for the inequality  $e(F; \Psi) \geq 1$  to hold for all  $\Psi$ . In particular if the logarithmic derivative of the density of  $F$  is bounded and satisfies additional regularity conditions, then for every  $\epsilon > 0$  there exists a true distribution  $\Psi$  such that  $e(F; \Psi) < \epsilon$ . If however  $\Psi$  differs from  $F$  only by location and scale parameters then  $e(F; \Psi) \geq 1$  with strict inequality holding unless either  $\Psi = F$  or  $F$  is normal.

**7. Group-Screening with More Than Two Stages.** M. S. PATEL, Research Triangle Institute.

The two stage group-screening procedure described earlier by W. S. Connor (Cf., The Proceedings of the Sixth Conference on the Design of Experiments in Army Research Development and Testing) and G. S. Watson (A study of the Group-screening method, August issue of *Technometrics*, 1961) is extended to multiple stages for the case when responses are observed with negligible error. Because of their potential usefulness, three and four stage procedures are treated in some detail. The general  $n + 1$  stage procedure is defined, a formula is developed for the expected number of runs, and for a fixed number of factors is minimized with respect to the sizes of the group-factors at various stages. Finally the procedures at different stages are compared with respect to the minimum expected number of runs.

(Abstracts of papers presented at the Eastern Regional Meeting of the Institute, New York City, December 27-30, 1961. Additional abstracts will appear in the March, 1962 issue.)

**1. Extensions of the Arc Sine Law.** SIMEON M. BERMAN, Columbia University.

An arc sine law for the number of positive partial sums in a sequence of "symmetrically dependent" random variables is obtained by means of the de Finetti representation theorem; this arc sine distribution is more general than that obtained by E. S. Andersen (1954) and holds under wider conditions. A secondary result of the paper is a direct generalization of an arc sine law of D. A. Darling (1951) for sums of independent random variables.

**2. Application of Simultaneous Confidence Intervals to Two Regression Problems (Preliminary report).** ARTHUR COHEN, Columbia University.

Consider the general linear hypothesis of full rank; that is, let  $y = Xb + u$  where  $y$  is an  $n \times 1$  vector of observations,  $X$  is a fixed  $n \times p$  matrix of rank  $p$ ,  $b$  is a  $p \times 1$  vector of parameters, and  $u$  is an  $n \times 1$  vector which is multivariate normal with mean vector zero and covariance matrix  $\sigma^2 I$ . H. Scheffé has shown how to obtain simultaneous confidence intervals for any number of estimable functions. His result is used to show how to obtain simultaneous confidence intervals for any number of parameters which are the ratios of linear combinations of the parameters in  $b$ . This latter result is applied to the multiple bioassay problem.

J. Mandel (*Ann. Math. Stat.*, Vol. 29 (1958), pp. 903-907) has shown how one might obtain simultaneous confidence intervals for any number of any real functions of the parameters in  $b$ . His result, along with the above mentioned result on the ratios of estimable functions, is used to test whether one quadratic regression function lies uniformly above another quadratic regression function over any given interval of abscissae.

**3. Contributions to the "Two-Armed Bandit" Problem.** DORIAN FELDMAN, Michigan State University.

The Bayes sequential design is obtained for an optimization problem involving the choice of experiments. Given are experiments  $A, B$ , densities  $p_1, p_2$ , a positive integer  $N$  (which may be  $\infty$ ) and a number  $\xi \in [0, 1]$ . A sequence of  $N$  observations is to be made such that at each stage either  $A$  or  $B$  is observed, the loss being 1 if the experiment with density  $p_2$  is

chosen, 0 otherwise.  $\xi$  is the prior probability that  $A$  has density  $p_1$  and the risk of a procedure is the expected number of observations on the experiment with density  $p_2$ , given  $\xi$ . Let  $R_N^A(\xi)$ ,  $R_N^B(\xi)$  denote the risks of the procedures that choose  $A$  first, respectively  $B$  first, and follow the optimal procedure for the last  $N - 1$  trials. It is shown inductively that for all  $N$  the difference between these risks is monotone in  $\xi$  and this is equivalent to optimality of the following procedure: At stage  $i + 1$  (regardless of  $N$ ) choose  $A$  or  $B$  according as  $\xi_i \geq \frac{1}{2}$  or  $\xi_i \leq \frac{1}{2}$ .  $\xi_i$  is the posterior probability that  $A$  has  $p_1$ . For  $N = \infty$ , the risk of this procedure is shown to be finite (hence optimal) and some specific risk functions are computed for binomial experiments.

**4. On the Axioms of Sample Formation and Their Bearing on the Construction of Linear Estimators in Sampling Theory for Finite Populations.** J. C. Koor, North Carolina State College.

Consider a universe (or population) of  $N$  elements described by a frame which in this case will be a simple list. In drawing a sample according to any probability system defined for the selection of the elements one at a time, either with or without replacement, three features inherent in the nature of the process of selection are evident. They are as follows: (i) the order of appearance of the elements, (ii) the presence or absence of any given element in the sample which is a member of the universe or population, and (iii) the set of elements composing the sample considered as one of the total number possible (in repeated sampling) according to the given probability system. As the veracity of the statements at (i), (ii) and (iii) is self-evident, they may be designated as axioms. It will be noted that the statements are not mutually contradictory. These features, inherent in the actual process of selection and as a result sample formation, supply the bases for the construction of linear estimators on a deductive basis. The axioms, which are implicit in the work of Horvitz and Thompson (*Ann. Math. Stat.*, Vol. 22 (1951), p. 315), considered singly, two at a time and most generally all three together, result in seven very general classes of linear estimators. The extension of the application of these axioms to samples from a universe with subdivisions (strata, first-stage units, second-stage units, etc.) is almost immediate.

**5. Maximum Likelihood Estimates for Certain Contagious Distributions Using High Speed Computers.** DONALD C. MARTIN AND S. K. KATTI, Florida State University.

Fortransit programs for fitting the Neyman Type A, the Poisson with Zeros and the simple Poisson, the Negative Binomial, and the Poisson Binomial by the method of maximum likelihood are available from the Statistics Department of the Florida State University. These programs have five operational modes allowing for combinations of the following: (i) Computing moment estimates. (ii) Using moment estimates or reading the initial parameter estimates and computing the maximum likelihood estimates by an iterative scheme. (iii) Reading in estimates computed by other means and using these to compute probabilities, expected frequencies and chi-square values, thus bypassing the maximum likelihood estimation process. (iv) Computing the probabilities, cumulative probabilities, expected frequencies, and term by term chi-square values and (v) Computing the chi-square value with some rudimentary grouping or the likelihood value. The chi-square section of the routine groups all expected frequencies less than a constant, usually 5, into a single cell. The Poisson is included as a special case of the Poisson with Zeros routine.

All programs are written in Fortransit II's for an IBM 650 computer with special characters. Running times vary widely between routines and data with the longest time on the order of 20 minutes and the shortest less than one minute. Typical times range from 1 to 10 minutes per distribution. An object program deck, a Fortransit statement deck, running

instructions and test data will be provided on request. All inquiries should be addressed to: The Department of Statistics, Florida State University, Tallahassee, Florida, attention S. K. Katti.

**6. Asymptotic Efficiency of a Class of  $c$ -Sample Tests.** M. L. PURI, University of California, Berkeley. (By title)

For testing the equality of  $c$  continuous probability distributions on the basis of  $c$  independent random samples, the test statistics of the form  $L = \sum_{j=1}^c m_j | (T_{N,j} - \mu_{N,j}) / A_N |^2$  are considered. Here  $m_j$  is the size of the  $j$ th sample,  $\mu_{N,j}$  and  $A_N$  are normalizing constants and  $T_{N,j} = \sum_{i=1}^{N_j} E_{N,i} Z_{N,i}^{(j)}$  where  $Z_{N,i}^{(j)} = 1$  if the  $i$ th smallest of  $N = \sum_{j=1}^c m_j$  observations is from the  $j$ th sample and  $Z_{N,i}^{(j)} = 0$  otherwise; and  $E_{N,i}$  are given numbers. Generalizing results of Chernoff-Savage (*Ann. Math. Stat.*, Vol. 29 (1958), pp. 972-994), sufficient conditions are given for the joint asymptotic normality of  $T_{N,j}$ ;  $j = 1, \dots, c$ . Under suitable regularity conditions and the assumption that the  $i$ th distribution function is

$$F(x + (\theta_i/N^{1/2}))$$

it is shown that as  $N \rightarrow \infty$ , the statistic  $L$  has a limiting noncentral chi-squared distribution. The asymptotic relative efficiency in Pitman's sense of the  $L$  test and the Kruskal-Wallis  $H$  test (which is a particular case of the  $L$  test) is obtained and is shown to be independent of  $c$ .

**7. Some Statistical and Operational Techniques in Reliability Studies (Preliminary report).** A. S. QUREISHI, IBM Service Bureau Corp., San Jose, California.

Given two production processes, the units from which fail in accordance with the Weibull Distribution, the problem is to select the particular process with the Smallest failure rate. It is assumed that there is a common guarantee period (known or unknown) during which no failures occur. It is desired to accomplish the above goal in as short a time as possible, thus maximizing the total gain without invalidating certain predetermined probability specifications. Three techniques, as suggested by Sobel (*Bell System*, Vol. 35 (1956) pp. 179-202) are considered and three procedures are constructed to show their advantages and disadvantages. Sobel's results on the assumption of exponential distribution turn out to be a particular case of the general solution presented in this paper. Expressions for average experiment time required to terminate the experiment have been obtained and the efficiencies of different procedures are compared. An alternative sequential procedure has been proposed and shown to be valuable in some situations. It has been shown with the aid of the tables (computed on Burroughs 220), that in some situations, one saves time and money if he assumes Weibull Distribution. Techniques have been developed to find the optimal sample size for each of the four procedures. Minimum Regret Principle has been applied to select the best procedure. The author is highly indebted to Professor N. L. Johnson, under whose Supervision this research is being carried.

**8. Asymptotic Bounds for the Zero-Crossing Probability Distribution of Stationary Gaussian Processes.** M. ROSENBLATT, Brown University.

Let  $X(t)$  be a separable stationary Gaussian process with mean zero and covariance function  $r(t) = E[X(\tau)X(\tau+t)]$ . Let  $G(T) = P[X(t) > 0, 0 \leq t \leq T]$ . Assume that  $r(t) \rightarrow 0$  as  $t \rightarrow \infty$ . It is then shown that  $G(T)$  approaches zero faster than any inverse power of  $T$  as  $T \rightarrow \infty$ . Stronger bounds are obtained for specific rates of decay of  $r(t)$  such as  $r(t) \sim t^{-\alpha}$ ,  $\alpha > 0$ , as  $t \rightarrow \infty$ . The basic tool is a powerful inequality of D. Slepian.

(Abstracts of papers not presented at any meeting of the Institute)

# 1. A Contribution to the Sphere-Packing Problem of Communication Theory.

A. V. BALAKRISHNAN, University of California, Los Angeles. (By title)

It is shown that the "sphere packing" problem (optimal band-limited signal selection for coherent Gaussian channels) can be reduced to the following extremal problem: "Given an  $N$ -variate Gaussian,  $\xi_1, \dots, \xi_N$ , with zero means and unit variances, maximize  $E(e^{\lambda \max_i \xi_i})$ , for  $\lambda > 0$ , with or without restriction on the rank  $m$  of the covariance matrix." It is shown that for a given  $m$  if there is a solution independent of  $\lambda$ , then this is given by maximizing the mean-width of the polyhedron generated by  $N$  unit vectors in Euclidean  $m$ -space. With no restriction on  $m$ ,  $m \leq N$ , it is shown that if there is a nonzero  $\lambda$ -interval for which the solution is independent of  $\lambda$ , this solution is given by the regular simplex in  $(N - 1)$  dimensions. Additional results using generalized tetrachoric series are given for the general problem.

# 2. Some Aspects of Statistical Invariance. DAVID R. BRILLINGER, Bell Telephone

Labs, Murray Hill, N. J. (By title)

Necessary and sufficient conditions are presented for a statistical problem to be invariant under a Lie transformation group. For the case of a (multi-dimensional) real-valued random variable  $x$  with c.d.f.  $F(x, \theta)$  the conditions reduce to:

(i) there must exist analytic functions  $\psi_i^\alpha(\theta)$ , or  $\sigma_i^\gamma(x)$  such that,

$$X_i F = \sum_{\alpha} \psi_i^\alpha(\theta) \frac{\partial F}{\partial \theta^\alpha} + \sum_{\gamma} \sigma_i^\gamma(x) \frac{\partial F}{\partial x^\gamma} = 0, \quad \text{for all } i$$

and

(ii) the infinitesimal generators  $X_i$  generate a group. The paper continues with a theorem concerning the distribution of statistics that are the maximal invariants of a compact topological transformation group. The theorem generalizes the technique that James has been using to derive some multi-dimensional distributions. The paper concludes with the following theorem justifying the "re-use" of samples: consider a random variable  $x$  with probability measure  $P$ . Let  $G$  be a set of measure preserving transformations of  $P$ . Let  $\varphi(x)$  be an unbiased estimator of  $\alpha$ , then  $\int_G \varphi(gx) d\mu(g)$  is also an unbiased estimator of  $\alpha$  and has smaller loss for any real valued convex loss function where  $\mu(g)$  is any measure of total mass 1 on  $G$ .

# 3. The $\delta$ -Method for Banach Valued Random Variables. DAVID R. BRILLINGER,

Bell Telephone Labs, Murray Hill, N. J. (By title)

The " $\delta$ -method" or method of "propagation of error" is extended so that it may be used in deriving the asymptotic distribution of Banach valued functions of Banach valued random variables. Define  $\text{plim } x_n = \theta$ ,  $\theta$  the 0 element of the Banach space, if for every  $\epsilon > 0$   $\lim P(\|x_n\| \leq \epsilon) = 1$ . Define  $\pi \lim x_n = \theta$ , if for every  $\epsilon > 0$  there exists an  $A_\epsilon$  such that  $P(\|x_n\| \leq A_\epsilon) \geq 1 - \epsilon$  for all  $n$ . Theorems proved include:

(1) Let  $x_n, y_n$  be sequences of Banach valued random variables with associated measures  $\mu_n, \nu_n$  on  $X$ . Let  $\{\nu_n\}$  converge weakly to a probability measure  $\mu$ . If  $\text{plim } (x_n - y_n) = \theta$ , then  $\{\mu_n\}$  converges weakly to  $\mu$ .

(2) Let  $x_n, x$  induce measures  $\mu_n, \mu$  on  $X$ . Let  $\mu_n$  converge weakly to  $\mu$ . Let  $g: X \rightarrow Y$  be a continuous map of  $X$  onto the Banach space  $Y$ . This map induces measures  $\{\nu_n\}, \nu$  such that  $\nu_n$  converges weakly to  $\nu$ .

(3) Let  $\{\lambda_n\}$  be a sequence of scalars such that  $|\lambda_n| \rightarrow \infty$ . Let  $g: X \rightarrow Y$  possess a Frechet differential everywhere in  $X$ . If  $\lim \lambda_n(x_n - y_n) = \theta$  then  $\text{plim } \lambda_n[g(x_n) - g(y_n) - dG(y_n, h_n)] = \theta$  where  $h_n = x_n - y_n$ .

**4. Some Fiducial Examples.** DAVID R. BRILLINGER, Bell Telephone Labs, Murray Hill, N. J. (By title)

This paper presents three examples with fiducial relevance. The first example concerns the definition of joint fiducial distributions. Quotes are given from works of Fisher concerning the genuine fiducial argument. An example is given that appears to satisfy all of Fisher's requirements, but yet that doesn't lead to a unique fiducial distribution. The second example demonstrates that recent results of Lindley concerning the identity of a fiducial distribution and a Bayes' *posteriori* distribution cannot be extended to higher dimensions in the obvious manner. The third example is the following: let  $x$  be  $N(\mu, 1)$ . The fiducial distribution of  $\mu^2$  may be derived in two manners, firstly from the fiducial distribution of  $\mu$  by means of a Jacobian multiplier, secondly from the fact that  $x^2$  is non-central  $\chi^2$  with parameter  $\mu^2$ . The results are different.



## NEWS AND NOTICES

*Readers are invited to submit to the Secretary of the Institute news items of interest*

### Personal Items

Richard Lee Beatty, Assistant Professor at the University of Wyoming, is presently on leave for a year of study at Stanford.

William H. Beyer completed his Ph.D. in Statistics at Virginia Polytechnic Institute in July, 1961, whereupon he joined the staff of the Department of Mathematics at the University of Akron (Ohio). He was formerly a member of the staff of the Department of Mathematics, V.P.I.

B. Ramdad Bhat has completed his Ph.D. in Statistics at the University of California, Berkeley and will be spending the academic year 1961-62 at Michigan State University as Assistant Professor of Statistics. Dr. Bhat is on leave from Karnatak University, Dharwar, India.

Professor R. C. Bose of the Department of Statistics at the University of North Carolina, has been granted leave to carry on research in Europe during the academic year 1961-62.

Mr. Victor Chew has accepted a part time position in the Department of Biostatistics of the Johns Hopkins University. He will divide his time between the Dr. David Duncan research project and the Operations Research Branch of the U. S. Naval Weapons Laboratory at Dahlgren, Virginia, where he is a mathematical statistician.

Dr. Edwin L. Crow, Consultant in Statistics at the Boulder Laboratories, National Bureau of Standards, is studying during the 1961-62 academic year at the Department of Statistics, University College, Gower Street, London, W.C.1, England.

Shanti S. Gupta, Member of the Technical Staff of the Bell Telephone Laboratories and Adjunct Professor of Mathematics at New York University, is on leave during the academic year 1961-62 and has been appointed Visiting Associate Professor in the Department of Statistics at Stanford University.

Francisco Azorín Poch, formerly of the Universidad Central de Venezuela teaching staff, has been appointed Professor of the Faculty of Science at the University of Santiago, Spain, beginning October 1st.

A. Salem Qureishi has joined the staff of IBM's subsidiary, San Jose Data Processing Center, as a statistician.

D. Raghavarao has completed his Ph.D. at the Bombay University.

Dr. Harry G. Romig, previously senior scientist and staff member of Operations Research, Inc., has been named corporate director of quality engineering of Leach Corporation.

Professor Dudley E. South, University of Florida, has accepted an appointment as visiting Professor of Mathematics, Florida Presbyterian College, St. Petersburg, for the academic year 1961-62.

### New Members

*The following persons have been elected to membership in the Institute*

- Armstrong, Robert K.**, B.A. (Boston University); Quality Engineer, Minneapolis/Honeywell, Commercial Division. *5th and 5th North, Minneapolis, Minnesota.*
- Bakhit, Osman B.**, B.Sc. (University of Kartoum, India); Graduate Student, Biometrics Unit, School of Agriculture, Cornell University, Ithaca, New York; *224 Linden Avenue, Ithaca, New York.*
- Blachman, Nelson M.**, Ph.D. (Harvard University); Consultant on Communication Theory, Electronic Defense Laboratories, Sylvania Electric Products, Inc., Box 205, Mountain View, California; *443 Ferne Avenue, Palo Alto, California.*
- Bohider, Neeti R.**, Ph.D. (Iowa State University); Assistant Professor of Statistics, *Utah State University, Logan, Utah.*
- Chanda, Kamal C.**, Ph.D. (University of Manchester, England); Assistant Professor of Mathematics, *University of Idaho, Moscow, Idaho.*
- Elliott, Antony G. L.**, B.Sc. (University of Western Australia); Lecturer in Mathematical Statistics, University of New South Wales; *Post Office Box 1, Kensington, NSW.; 1 Yanko Road, Pymble, NSW.*
- Farrell, Edward J.**, B.S., (University of Minnesota); Student at University of Minnesota; Remington Rand Univac, Univac Park, St. Paul 16, Minnesota; *1370 Berkeley Avenue, St. Paul 6, Minnesota.*
- Foster, Louis A.**, M.S., (Purdue University); Graduate Student, *Statistical Laboratory, Purdue University, Lafayette, Indiana.*
- Giri, Narayan C.**, Ph.D. (Stanford University); Teaching Assistant, *Statistics Department, Stanford University, Stanford, California.*
- Glass, Donald N.**, B.S., (University of Wyoming); Chief of Computing, Computing Laboratory, University of Wyoming, Department of Statistics; *463 North 10th Street, Laramie, Wyoming.*
- Godfrey, Milton L.**, B.M.E., (New York University); Director of Applied Sciences Division, *CIEB, 270 Park Avenue, New York 17, New York.*
- Goldfab, Jay**, M.B.A., (University of Pennsylvania); Operations Research Analyst, *Johnsville, Pennsylvania.*
- Gopalan, M. N.**, M.Sc., (University of Mysore, India); Senior Technical Analyst, Indian Institute of Technology, Department of Mathematics, *Post Office Box 11T, Powai, Bombay 76, India.*
- Gray, Kenneth B. Jr.**, M.S., (Stanford University); Graduate Research Mathematician, U.C.L.A., Los Angeles 24, California; *3100 Sawtelle Boulevard, Los Angeles 66, California.*
- Kern, Richard N.**, M.S. (St. Louis University); Supervisory Mathematician, National Security Agency, Fort George Meade; *904 Kenbrook Court, Silver Spring, Maryland.*
- Kinney, John J.**, M.S. (University of Michigan); Assistant Professor of Mathematics, St. Lawrence University, Canton, New York; *55 Perine Street, Dansville, New York.*
- Klebba, A. Joan**, M.A. (Catholic University of America); Staff Statistician, National Office of Vital Statistics; *1489 South 28th Street, Arlington, 6, Virginia.*
- Lowenstein, Regina S.**, A.M., (Columbia University); Statistician, Research Unit, Columbia University School of Public Health and Administrative Medicine, 21 Audubon Avenue, New York 33, New York; *33 West 63rd Street, New York 23, New York.*
- Nederlof, Marinus Herman**, Ph.D., (Leiden State University, The Netherlands); Office Geologist for Compania Petrolera Boliviana Shell Ltda., Casilla de Correos 20, La Paz, Bolivia; *c/o Ir. O. B. Bloasma, Graaf Jan laan 11, Naarden, The Netherlands.*
- Pate, James L.**, M.A. (University of Alabama); Instructor, University of Alabama; *Box 4211, University, Alabama.*
- Pessin, Vivian**, M.A. (Columbia University); Statistician, Childrens Hospital, *219 Bryant Street, Buffalo 22, New York.*

- Peugh, Laura V.**, M.S. (Purdue University); Mathematician, *David Toller Model Basin, Carderock, Maryland.*
- Powell, Joe L.**, M.S. (Tulane University); Assistant Director, Biomathematics Laboratory, Baylor University College of Medicine, Texas Medical Center, Houston, Texas; *8245 Park Place Boulevard, Apt. 6, Houston 17, Texas.*
- Reyment, Richard A.**, D.Sc. (University of Stockholm); Docent at the *University of Stockholm, Department of Geology, Kungstengatan 45, Stockholm Va., Sweden.*
- Rhode, Charles A.**, B.S. (Case Institute of Technology, Cleveland, Ohio); Graduate Assistant, *Institute of Statistics, North Carolina State College, Raleigh, North Carolina.*
- Rowe, Peter M.**, M.B.A., (New York University); Student, New York University Graduate School of Arts and Sciences, Washington Square, New York; *211 Alpine Drive, Closter, New Jersey.*
- Sarndal, Carl-Erik**, Fil.Lic. (University of Lund, Sweden); Post-doctoral fellow, *Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina.*
- Schafer, Ray E.**, M.B.A. (Western Reserve, Cleveland, Ohio); MTS Statistician, Hughes Aircraft Corporation, Ground Systems Group, Box 2097, Fullerton, California, Mail Station E 116, Building 393; *11662 Stuart Drive, Garden Grove, California.*
- Sellheim, H. Dale**, B.S., (University of North Dakota, Grand Forks); Senior Mathematician, C-E-I-R, 1200 Jefferson Davis Highway, Arlington 2, Virginia; *6713 14th Street N.W., Washington, D.C.*
- Seri, Armand**, M.S. (University of Wichita); Graduate Student, University of Illinois, Urbana; *905 South First Street, Apt. 24, Champaign, Illinois.*
- Tantuco, Guillermo N.**, M.A., (University of Michigan); *Tantuco Enterprises, 311 Ayala Building, Juan Luna Street, Manila, Philippines.*
- Tierney, Thomas R.**, B.S., (Temple University); Student; *1926 North Broad Street, Philadelphia 21, Pennsylvania.*
- Whitton, Howard J. G.**, B.Sc., (University of London); Senior Mathematician, Trans Canada Airlines (O.R. Department) Room 907, Terminal Centre Building, 1060 University Street, Montreal 3, Quebec, Canada; *2407 Lockhart Avenue, Montreal 16, Quebec, Canada.*
- Williams, Bryan Mc.**, A.B., (George Washington University); Booz/Allen Applied Research, Inc., 4921 Auburn Avenue, Bethesda, Maryland; *4611 River Road, N.W., Washington 1, D. C.,*
- Winter, Ralph P.**, M.Ed., (College of St. Thomas, St. Paul, Minnesota); Graduate Student, University of Minnesota, Minneapolis 14; *5712 38th Avenue South, Minneapolis 17, Minnesota.*

---

### INTERNATIONAL ABSTRACTS IN OPERATIONS RESEARCH

The Operations Research Society of America and the International Federation of Operations Research Society announce the publication of a new journal, *International Abstracts in Operations Research*. Herbert P. Galliher (Massachusetts Institute of Technology) will be the Editor.

The new journal will contain abstracts of articles selected, on the basis of their relationship to Operations Research, from a large number of journals. The first issue is scheduled to appear in November, 1961. An annual subscription will cost \$12.50 in the United States and Canada, \$10 elsewhere.

Orders for subscriptions should be sent to the Business Manager, Operations Research Society of America, Mt. Royal and Guilford Avenues, Baltimore 2, Maryland, U.S.A. Editorial communications should be addressed to the Editor,

International Abstracts in Operations Research, Room 6-218, Massachusetts Institute of Technology, Cambridge 39, Mass., U.S.A.

---

### FELLOWSHIP AND RESEARCH OPPORTUNITIES

The Division of Mathematics, National Academy of Sciences—National Research Council calls attention to a variety of fellowships and other support for basic research in mathematics to be awarded by agencies of the Federal Government during the year 1961–62. A list of sources of support is given in the bulletin, "A Selected List of Major Fellowship Opportunities and Publications for Educational Support," available from the Fellowship Office, National Academy of Sciences—National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

---

### PRELIMINARY ACTUARIAL EXAMINATIONS PRIZE AWARDS ANNOUNCED

The winners of the prize awards offered by the Society of Actuaries to the nine undergraduates ranking highest on the score of the General Mathematics Examination of the 1961 Preliminary Actuarial Examinations are as follows:

#### *First Prize of \$200*

Wells, John C.

Massachusetts Institute of Technology

#### *Additional Prizes of \$100 each*

Bender, Edward A.

California Institute of Technology

Butler, William A.

Queen's University

Corwin, Lawrence J.

Harvard University

Hochster, Melvin

Harvard University

Mather, John N.

Harvard University

Segal, David M.

Harvard University

Waterhouse, William C.

Harvard University

Weiss, Norman J.

Harvard University

The Society of Actuaries has authorized a similar set of nine prizes for 1962. The Preliminary Actuarial Examinations consists of two examinations: The General Mathematics Examination (based on the first two years of college mathematics), and The Probability and Statistics Examination. The 1962 Preliminary Actuarial Examinations will be prepared by the Educational Testing Service under the direction of a committee of actuaries and mathematicians, and will be administered by the Society of Actuaries at centers throughout the United States and Canada on November 15, 1961 and on May 9, 1962. The closing date for applications is April 1, 1962. Further information concerning these Examinations can be obtained from the Society of Actuaries, 208 South LaSalle Street, Chicago 4, Illinois.

### DISCOUNT FOR THEORY OF PROBABILITY AND ITS APPLICATIONS

The Society for Industrial and Applied Mathematics has announced that it is extending its discount subscription rate for *Theory of Probability and its Applications* (a cover-to-cover English translation of the Soviet journal, *Teoriya Veroyatnostei i ee Primeneniya*) to the individual members of the organizations comprising the Conference Board of the Mathematical Sciences. In particular, members of the Institute of Mathematical Statistics may now subscribe at the rate of \$9.50 a year; subscriptions may be entered beginning with Volume 1 (1956) or beginning with later volumes.

Volumes 1 (1956) through 4 (1959) are now complete and available. The first three of these are bound in hard covers. All the 1960 issues should be available by the end of 1961.

Subscription orders, and requests for further information, should be sent to SIAM, Box 7541, Philadelphia 1, Pennsylvania, U. S. A. Membership in the Institute of Mathematical Statistics should be mentioned.

---

### SYMPOSIUM ON REDUNDANCY TECHNIQUES FOR COMPUTING SYSTEMS

A symposium on Redundancy Techniques for Computing Systems, sponsored by the Information Systems Branch, Office of Naval Research, will be held on 6 and 7 February 1962. This Symposium will be held in the *Department of the Interior Auditorium*, C Street between 18th and 19th Street, N. W., Washington, D. C.

The objective of the Symposium is to focus attention on new ideas, research, and developments which may lead to the introduction of redundancy techniques into forthcoming computing systems. It is apparent that many information processing systems of the future will be of such size that conventional methods of fabrication and of maintenance will be completely infeasible. Computers for space applications have reliability requirements which are presently not achievable without use of redundancy. Thus, some form of logical redundancy must be introduced if practical systems are to be operated. Accordingly, it is planned to present in a single symposium a collection of related papers concerning suggested new techniques for the use of redundancy in computing systems. This should permit relative evaluation of such techniques with respect to their probable future utilization in various aspects of the computing field. The program will consist of papers invited from many of the organizations engaged in appropriate research and development activities.

Attendance is open to all interested technical personnel. Further information and a preliminary Symposium program, when available, may be obtained from Miss Josephine Leno, Code 430A, Office of Naval Research, Washington 25, D. C., (Phone OXford 6-6213).



## REPORT OF THE PRESIDENT FOR 1961

The past year has been one of continued growth for the Institute as for the statistical profession as a whole.

*Publications*

As announced in the March *Annals*, the year saw the publication of the first two volumes of the *Statistical Research Monographs*, jointly sponsored by the Institute and the University of Chicago. Another new series, of which the first volume was published this year, is the *Selected Translations in Mathematical Statistics and Probability*, which is being published jointly by the Institute and the American Mathematical Society. Further volumes in both series are in preparation.

The most central activity of the Institute is probably its publication of the *Annals of Mathematical Statistics*. The *Annals* was founded in 1930 by Professor H. C. Carver, and he edited the journal—which was published first under the auspices of the American Statistical Association and later of the IMS—until 1938. In recognition of the great debt we all owe to Professor Carver for this farsighted and courageous action which he undertook against great odds and discouragement, and as a token of our gratitude to him, the Council is dedicating the 1961 volume of the *Annals* to Professor Carver on the occasion of his retirement from his Professorship at the University of Michigan.

The end of the Institute year marks the expiration of the term of office of the Editorial Board of the *Annals*. I wish to express the gratitude of the Institute to this board and in particular to the retiring Editor, William Kruskal, for his untiring and most successful work during the last three years. At the recommendation of a committee consisting of L. J. Savage (Chairman), D. R. Cox, T. E. Harris and C. M. Stein, the Council asked J. L. Hodges, Jr., to take over the Editorship and Professor Hodges has accepted this appointment.

The Council decided at the Stanford meeting that it would be desirable to review the organization of the *Annals*, and in particular to investigate whether the Institute should start a separate probability journal. Much thought was given to these problems by a committee consisting of T. E. Harris (Chairman), D. Blackwell, J. Daly, J. L. Doob, J. L. Hodges, Jr., W. Kruskal and W. Smith. The committee recommended against the establishment of a separate probability journal. Regarding the organization of the *Annals*, it pointed to the imperative need to cut down the work load of the Editor, and to that end recommended that the Editor should be freed of the duties connected with publishing (as opposed to editing) the *Annals*. The Committee also suggested that the whole problem of *Annals* reorganization should be considered again before the end of the term of the incoming Editor.

*Meetings*

An important feature of the meetings held this year, a spring meeting of the Eastern Region at Cornell and the Annual meeting in Seattle, were a number of Special Lectures arranged by the Committee on Special Lectures, consisting of



Leo Katz (Chairman), G. E. P. Box, D. Gilford, J. Kiefer, H. Robbins, H. Scheffé, and D. Wallace. The 8th Rietz lecture was delivered by Professor Blackwell on "Dynamic Programming" and the 4th series of Wald lectures by Professor Stein on the topic "Estimation of Many Parameters." In addition, Special Invited Papers were presented by Dr. T. Dalenius, Professor D. Dugué and Professor M. Fisz.

At the Stanford meeting, the Council decided to initiate the holding of European regional meetings in addition to the Annual meeting and to the meetings of the American regions. At the recommendation of a Committee consisting of H. Theil (Chairman), J. Durbin, G. Elfving, M. Fisz, R. Fortet, U. Grenander, A. Hald, J. Hemelrijk, A. Linder, D. Lindley, L. Schmetterer and E. Sverdrup, the Council at the Seattle meeting accepted invitations to hold such meetings in the summer of 1962 in Dublin and of 1963 in Copenhagen.

After an interruption of several years, we hope to resume the Summer Institutes which were so successful in 1955-58. At the recommendation of a committee consisting of J. Kiefer (Chairman), H. Chernoff, D. Gilford, J. Pratt and M. Sobel, the Council accepted the invitation of Michigan State University to hold a Summer Institute there in 1963 on the subject, "Statistical inference in stochastic processes"; the committee hopes also to make arrangements for such an Institute for 1962.

#### *Membership*

The Committee on Institutional Memberships under the chairmanship of Mervin Muller has had another very successful year. President-Elect Bowker organized a drive for individual members which also met with considerable success. Finally, a European Membership Committee has just started functioning under the chairmanship of Eugene Lukacs.

It is a great pleasure to congratulate the new Fellows of the Institute: R. Bradley, T. Dalenius, C. Derman, M. Dwass, D. Gilford, M. V. Johns, Jr., E. Parzen, and I. R. Savage.

#### *Visiting Scientist's Program*

Because of the critical shortage, in many areas, of qualified statisticians, the Council at the Stanford meeting discussed the desirability of a Visiting Scientist's Program in Statistics and Probability, which would bring the nature of the field and its potentialities to the attention of prospective students and their advisers. A proposal to the National Science Foundation for support of such a program has been prepared by a committee consisting of R. Bradley (Chairman), D. Blackwell, D. Chapman, J. W. Hamblen, B. Harshbarger, R. Pinkham, L. Snell and S. S. Wilks.

#### *The Role of IMS and Its Relationship with Other Societies*

Although the IMS was founded in the United States and most of its activities have been centered there, the constitution has never provided for a national affiliation. At present, about one fourth of the IMS membership is outside the U.S.A., and the proportion has been increasing.

With the rapid development of statistics in many parts of the world, there is an increasing need for a nonrestrictive and truly international statistical organization, with emphasis on theory. It is possible that, in time, the IMS could develop into such an organization. The whole problem of the international role of IMS, including in particular the question of whether IMS should affiliate with the International Statistical Institute, is being studied by a committee consisting of W. G. Cochran (Chairman), M. S. Bartlett, J. Durbin, D. M. Gilford, U. Grenander, A. Hald, J. Neyman, G. E. Nicholson and J. W. Tukey.

In the U.S.A., the growth of statistics has posed the problem of developing an organization that could effectively speak for the statistical community as a whole. For this purpose, the American Statistical Association organized a meeting of statistical societies, at which our representatives were H. Scheffé and S. S. Wilks. At the meeting, a conference board of statistical societies was discussed, similar to the Conference Board of Mathematical Sciences of which IMS is a member, and a committee was appointed to draft a constitution for such a board.

A great deal of work of the Institute is done by its committees, and I should like to thank the many persons who accepted committee service and worked hard to make it a success. I should also like to acknowledge the great personal debt I owe to President-Elect Albert H. Bowker and Program Coordinator Dorothy Gilford for their invaluable advice and assistance throughout the year. In conclusion I should like to announce the Nominating Committee for next year, which consists of H. Raiffa (Chairman), H. E. Daniels, P. Meier, D. Owen, R. Pyke, C. R. Rao, A. Rényi, I. R. Savage, and M. Zelen.

E. L. LEHMANN

### IMS OFFICERS, COMMITTEES, AND REPRESENTATIVES—1961

#### *Council Members and Officers*

##### *Terms Expire 1961*

F. J. Anscombe  
T. E. Harris  
Leo Katz  
S. S. Wilks

##### *Terms Expire 1963*

Herman Chernoff  
Kai-Lai Chung  
M. G. Kendall  
Gerald J. Lieberman  
Charles Stein

##### *Terms Expire 1962*

T. W. Anderson  
Z. W. Birnbaum  
J. L. Hodges, Jr.  
Wassily Hoeffding

##### *Terms Expire 1964*

R. L. Anderson  
D. Blackwell  
W. Kruskal  
S. N. Roy

##### *One Year Terms*

Joseph Berkson  
David L. Wallace

*President:* E. L. Lehmann  
*President-Elect:* A. H. Bowker  
*Secretary:* G. E. Nicholson, Jr.  
*Treasurer:* Gerald Lieberman  
*Editor:* W. H. Kruskal

*Program Coordinator:* D. H. Gilford  
*Associate Secretaries*  
*Central:* D. Burkholder  
*Eastern:* Joan Rosenblatt  
*Western:* Fred Andrews

## COMMITTEES—1961

*(The first person named is the chairman)*

## COMMITTEE ON ANNALS INDEX

I. R. Savage, T. E. Harris, J. L. Hodges, Jr., W. H. Kruskal, G. E. Nicholson, Jr.

## BROCHURE COMMITTEE

E. Parzen, D. Blackwell, A. H. Bowker, J. F. Daly, B. G. Greenberg, M. H. Hansen,  
G. E. Nicholson, Jr., S. S. Wilks

## COMMITTEE ON EXCHANGES

Work of committee now performed by G. E. Nicholson, Jr.

## COMMITTEE ON FELLOWS

H. Solomon (1961), E. J. G. Pitman (1961), W. G. Cochran (1962), J. L. Hodges, Jr.  
(1962), G. Elfving (1963), I. Olkin (1963)

## COMMITTEE ON FINANCE

H. Levene, K. J. Arnold, K. A. Brownlee, J. Curtiss, G. Lieberman

## COMMITTEE ON INSTITUTIONAL MEMBERS

M. E. Muller, K. J. Arnold, E. L. Crow, M. Dwass, E. Lukacs, P. D. Minton,  
R. B. Murphy, E. Seiden

## COMMITTEE ON MATHEMATICAL TABLES

D. B. Owen, R. L. Anderson, A. H. Bowker, P. C. Cox, W. J. Dixon, C. Eisenhart,  
J. A. Greenwood, S. S. Gupta, H. O. Hartley, H. L. Harter, L. Katz, F. C.  
Leone, M. E. Muller, P. Olmstead, G. P. Steck, M. A. Woodbury, M. Zelen

## COMMITTEE ON MEMBERSHIP

A. H. Bowker

## NOMINATING COMMITTEE—1962 ELECTION

Howard Raiffa, H. E. Daniels, P. Meier, D. B. Owen, R. Pyke, C. R. Rao,  
A. Rényi, I. R. Savage, M. Zelen

## COMMITTEE ON PROFESSIONAL STANDARDS

J. Lev, R. W. Burgess, C. Eisenhart, G. M. Harrington, B. Kimball,  
A. W. Kimball, H. Marshall, R. Patton, J. Walsh

## PROGRAM COMMITTEE FOR ANNUAL MEETING

D. Wallace, F. Anscombe (Vice-Chairman), J. Blum, J. L. Folks, R. Gnanadesikan,  
A. T. James, R. Mickey, R. Miller, R. Radner, G. Watson, G. E. Nicholson, Jr.,  
D. M. Gilford (ex officio)

## PROGRAM COMMITTEE FOR CENTRAL REGIONAL MEETING

R. A. Wijsman, J. Berkson, G. E. P. Box, V. R. Hogg, M. Katz, Jr., P. Minton,  
H. Rubin, M. Sobel, O. Wesler, D. R. Whitney

## PROGRAM COMMITTEE FOR EASTERN REGIONAL MEETING

Mervin E. Muller, R. Bechhofer, R. A. Bradley, C. Daniel, A. P. Dempster,  
S. W. Greenhouse, M. Kastenbaum, E. Paulson, W. L. Smith,  
H. Teicher, R. Wormleighton, M. Zelen

## PROGRAM COMMITTEE FOR EUROPEAN MEETING

H. Theil (Rotterdam), G. Elfving (Stanford), R. Fortet (Paris), M. Fisz (Seattle), U.  
Grenander (Stockholm), A. Hald (Copenhagen), J. Hemelrijk (Amsterdam), D. V.  
Lindley (Cardiganshire, Wales), A. Linder (Geneva), L. Schmetterer (Ham-  
burg), E. Sverdrup (Oslo)

## COMMITTEE ON THE INTERNATIONAL ROLE OF IMS

W. Cochran, M. S. Bartlett, J. Durbin, D. Gilford, U. Grenander,  
A. Hald, J. Neyman, G. Nicholson, J. W. Tukey

## COMMITTEE TO SELECT A NEW EDITOR

L. J. Savage, D. Cox, T. Harris, C. Stein

## COMMITTEE ON SPECIAL PAPERS

L. Katz, G. E. P. Box, J. Kiefer, I. Olkin, H. Robbins, H. Scheffé,  
D. Gilford (ex officio), D. Wallace (ex officio)

## COMMITTEE ON SUBSCRIPTIONS

E. P. Coleman, J. K. Adams, C. B. Bell, K. A. Bush, L. R. Elveback, H. Harmon

## COMMITTEE ON ANNALS REORGANIZATION

T. Harris, D. Blackwell, J. F. Daly, J. L. Doob, J. L. Hodges, Jr., W. H. Kruskal,  
W. Smith

COMMITTEE TO REEXAMINE IMS FROM VIEWPOINT OF  
YOUNGER MEMBERS

W. L. Smith, P. Billingsley, S. H. Brooks, B. W. Brown, Jr., D. L. Burkholder, M. B.  
Danford, M. H. DeGroot, A. P. Dempster, D. A. Gardiner, W. J. Hall, J. W. Hamblen,  
J. E. Jackson, A. Madansky, H. E. McKean, R. Pyke, D. L. Wallace, O.  
Wesler, R. A. Wijsman

## COMMITTEE ON RUSSIAN TRANSLATIONS

W. Hoeffding, M. Dwass, E. Lukacs, I. Olkin, L. Schmetterer, W. Kruskal (ex officio)

## COMMITTEE TO ORGANIZE A SUMMER INSTITUTE

J. Kiefer, H. Chernoff, J. Pratt, M. Sobel

## COMMITTEE ON THE VISITING SCIENTISTS PROGRAM

R. Bradley, D. Blackwell, D. Chapman, J. W. Hamblen, B. Harshbarger,  
R. Pinkham, L. Snell, S. S. Wilks

## COMMITTEE ON THE RELATIONSHIPS AMONG STATISTICAL SOCIETIES

H. Scheffé, S. S. Wilks

## REPRESENTATIVES TO PROFESSIONAL ASSOCIATIONS—1961

IMS REPRESENTATIVE TO AAAS—Harold Hotelling

AMERICAN STANDARDS ASSOCIATION COMMITTEE ON STATISTICAL  
NOMENCLATURE: IMS REPRESENTATIVE—P. G. Hoel

IMS REPRESENTATIVE TO CONFERENCE ORGANIZATION OF THE  
MATHEMATICAL SCIENCES—W. M. Rosenblatt, Z. W. Birnbaum

IMS REPRESENTATIVE IN DIVISION OF MATHEMATICS—NATIONAL  
RESEARCH COUNCIL—F. C. Mosteller

IMS REPRESENTATIVES ON JOINT ASA-IMS BROCHURE COMMITTEE

S. S. Wilks, D. Blackwell, B. G. Greenberg, W. H. Kruskal

IMS REPRESENTATIVES TO AMS-IMS COMMITTEE ON RUSSIAN  
TRANSLATIONS—W. Hoeffding, I. Olkin

### REPORT OF THE EDITOR FOR 1961

During the operating year August 1, 1960 to July 31, 1961, there were 225 manuscripts submitted to the *Annals*. Final decisions were made for 219 manuscripts during the 1960-61 operating year, and 185 manuscripts were under editorial consideration on July 31, 1961.

A detailed statistical report of *Annals* operations in 1960-61 will be sent to interested members on request.

William Kruskal's term as Editor ended on June 30, 1961; J. L. Hodges, Jr. began his editorship on July 1, 1961. Kruskal will, however, continue to be responsible for editorial decisions about manuscripts submitted on or before June 30, 1961, and still under consideration at that date.

It is a pleasure to express renewed gratitude to the Associate Editors for their difficult and effective work. Mrs. Robert Isherwood, Editorial Assistant, has devoted herself to the complex task of managing and carrying out the many tasks that arise in publishing the *Annals*; we thank her deeply. The Universities of Chicago and California have continued to provide material aid.

Finally, we have the pleasure of listing the names of referees of papers for which final editorial decisions have been made during the period from October 1960 and September 1961 inclusive. Their work is essential in carrying out our editorial procedures, and we express to them our great appreciation.

WILLIAM KRUSKAL, Past Editor

J. L. HODGES, JR., Editor

Addelman, Sidney

Anderson, R. L.

Anderson, T. W.

Anscombe, F. J.

Armitage, P.

Arrow, Kenneth

Bahadur, R. R.

Balakrishnan, A. V.

Barlow, Richard E.

Baxter, Glen

Bharucha-Reid, A. T.

Billingsley, Patrick

Birnbaum, Allan

Blackwell, David

Blum, Julius R.

Blumenthal, R. M.

Bofinger, V.

Borges, Rudolf E.

Bose, R. C.

Box, G. E. P.

Brillinger, David

Brunk, H. D.

Buehler, Robert J.

Burkholder, D. L.

Chang, C. L.

Chakravarti, I. M.

Chapman, D. G.

Cogburn, Robert

Connor, William S.

Craswell, Keith

Daley, Joseph F.

Daniels, H. E.

Darling, Donald A.

Dean, B. V.

DeGroot, Morris

Dempster, A. P.

Derman, Cyrus

Dhruvarajan, P. S.

Dixon, W. J.

- |                      |                     |                      |
|----------------------|---------------------|----------------------|
| Duncan, David B.     | Lehmann, E. L.      | Savage, I. R.        |
| Dwass, Meyer         | Lieberman, Gerald   | Savage, L. J.        |
| Eicker, Friedhelm    | Lindley, Dennis V.  | Schaufele, Ronald    |
| Epstein, Ben         | Lloyd, S. P.        | Scheffé, Henry       |
| Fisz, M.             | Loève, Michel       | Schwarz, Gideon      |
| Folks, J. Leroy      | Lukacs, Eugene      | Shah, B. V.          |
| Foster, F. G.        | Madansky, A.        | Siegert, Arnold      |
| Fraser, D. A. S.     | Mallows, Colin L.   | Smith, Walter        |
| Ghurye, S. G.        | Marsaglia, G.       | Snell, J. Laurie     |
| Gnanadesikan, R.     | Marshall, Albert W. | Sobel, Milton        |
| Good, I. J.          | Mason, D. D.        | Spitzer, Frank       |
| Goodman, Leo A.      | Massey, Frank, Jr.  | Sprott, David A.     |
| Graybill, F.         | Mesner, Dale        | Steck, G. P.         |
| Grenander, Ulf       | Meyer, Paul         | Stein, Charles M.    |
| Hadley, George       | Miller, Rupert G.   | Takács, L.           |
| Hall, W. J.          | Moses, Lincoln      | Tate, R. F.          |
| Halperin, Max        | Mulholland, H. P.   | Teicher, Henry       |
| Hammersley, J. M.    | McCarthy, Philip    | Teichroew, D.        |
| Harter, H. Leon      | McKean, Harlley     | Thomasian, Aram J.   |
| Hastings, W. Keith   | Nash, Stanley       | Thompson, Donovan    |
| Herbach, Leon        | Noether, Gottfried  | Thompson, W. A., Jr. |
| Hill, Bruce M.       | Olkin, Ingram       | Trotter, Hale F.     |
| Hoeffding, Wassily   | Owen, Donald B.     | Truax, Donald R.     |
| Hogg, Robert V.      | Parzen, Emanuel     | Tucker, Howard       |
| Hook, Robert         | Patel, M. S.        | Tukey, John W.       |
| Hotelling, Harold    | Paulik, Gerald      | Tweedie, M. C. K.    |
| James, A. T.         | Paulson, E.         | Ury, Hans            |
| Jeeves, J. A.        | Plackett, R. L.     | Wallace, David L.    |
| Johnson, N. L.       | Pratt, John         | Watson, G. S.        |
| Jones, Richard       | Proschan, Frank     | Weiss, Lionel        |
| Kallianpur, Gopinath | Pyke, Ronald        | Welch, B. L.         |
| Katz, Melvin L.      | Rényi, A.           | Wendel, James        |
| Kemperman, J. H. B.  | Richter, Donald     | Wesler, Oscar        |
| Kempthorne, Oscar    | Robbins, Herbert    | Whittle, Peter       |
| Kiefer, Jack         | Robson, D. S.       | Wijsman, Robert      |
| Kruskal, Martin D.   | Rosenblatt, Murray  | Wilk, M. B.          |
| Kullback, S.         | Roy, S. N.          | Williams, J.         |
| Laha, R. G.          | Ruben, Harold       | Wolfowitz, J.        |
| Lamperti, John       | Ruben, Herman       | Ylvisaker, Donald    |
| Laurent, André       | Rustagi, J. S.      | Zelen, Marvin        |
| LeCam, Lucien        | Saunders, S.        |                      |

#### DOCTORAL DISSERTATIONS IN STATISTICS, 1960

The following dissertation titles were received too late for inclusion in the September 1961 *Annals*.

Ed.

**Martin Arthur Arkowitz**, Cornell University, major in geometry, algebra, analysis, "The Generalized Whitehead Product."

**K. M. Ferrin**, University of California, Los Angeles, major in statistics, "Multiple Decision Procedures for Normal Populations."



- Yoichiro Fukuda, University of California, Los Angeles, major in statistics, "Estimation Problems in Inventory Control."
- R. I. Jennrich, University of California, Los Angeles, major in statistics, "Analysis of Variance in the General Mixed Model."
- Alan Gustave Koneheim, Cornell University, major in analysis, algebra, applied mathematics, "Some Properties of a Class of Finite Trigonometric Sums."
- Samuel Kotz, Cornell University, major in analysis, algebra, geometry, "Exponential Bounds for the Probability of Error in Discrete Memoryless Channels."
- E. M. Scheuer, University of California, Los Angeles, major in statistics, "Simultaneous Estimation for Means of Medians of Dependent Random Variables without Distribution Assumptions."

---

### PUBLICATIONS RECEIVED

- Anuario Estadístico de España*, 1961 Edición Manual, Presidencia del Gobierno, Instituto Nacional de Estadística, Ferraz 41, Madrid, Spain, 1001 pp.
- Bedkenbach, Edwin and Richard Bellman, *An Introduction to Inequalities*, Random House, New York, 1961, paperback, 133 pp. \$1.95.
- Davis, Philip J., *The Lore of Large Numbers*, Random House, New York, 1961, paperback, 165 pp. \$1.95.
- Friend, J. Newton, *More Numbers: Fun and Facts*, Charles Scribner's Sons, New York, 1961, 201 pp. \$2.95.
- Gersdorff, Ralph von, *Portugals Finanzen: Geschichtlicher Überblick Die Finanzreformen Prof Salazars Steuer- und Staatsschuldenwesen*, Verlag Ernst und Werner Giesecking, Bielfeld, 1961, 280 pp.
- Hidano, Naoru and Masatoshi Sedani, *Statistics for Psychology and Education* (in Japanese), Baifukan Co., Tokyo, Japan, 1961, 346 pp.
- "Investing in Scientific Progress, 1961-1970: Concepts, goals, and projections," National Science Foundation, Washington, D. C., 1961, 30 pp.
- Kazarinoff, Nicholas D., *Geometric Inequalities*, Random House, New York, 1961, paperback, 132 pp. \$1.95.
- Mosteller, Frederick, Robert E. K. Rourke, and George B. Thomas, Jr., *Probability: A First Course*, Addison-Wesley Pub. Co., Reading, Mass., 1961, 319 pp. \$5.00.
- Mosteller, Frederick, Robert E. K. Rourke, and George B. Thomas, Jr., *Probability With Statistical Applications*, Addison-Wesley Pub. Co., Reading, Mass., 1961, 478 pp. \$6.50.
- Murray, Francis J., *Mathematical Machines, Vol. I: Digital Computers*, Columbia University Press, New York, 1961, 300 pp. \$12.50.
- Murray, Francis J., *Mathematical Machines, Vol. II: Analog Devices*, Columbia University Press, New York, 1961, 365 pp. \$17.50.
- Niven, Ivan, *Numbers: Rational and Irrational*, Random House, New York, 1961, paperback, 136 pp. \$1.95.
- Rasch, G., *Probabilistic Models for Some Intelligence and Attainment Tests*, Studies in Mathematical Psychology I, Danmarks Paedagogiske Institut, Denmark, 1960.
- Salkind, Charles T., Ed., *The Contest Problem Book*, Random House, New York, 1961, paperback, 152 pp. \$1.95.
- Sawyer, W. W., *What is Calculus About?* Random House, New York, 1961, paperback, 118 pp. \$1.95.



ADVERTISING IN  
THE ANNALS of  
MATHEMATICAL STATISTICS

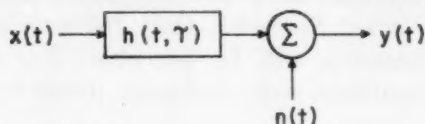
ADVERTISEMENTS for books, recruitment of professional personnel, etc., may now be placed in the *Annals of Mathematical Statistics*. Only full-page and half-page advertisements will be accepted. For details about costs, deadlines, sizes, and so on, please write to

Mr. Edgar M. Bisgyer  
Advertising Manager  
American Statistical Assn.  
1757 K Street, N.W.  
Washington 6, D. C.

---

## DECISION AND THE RANDOM CHANNEL

For some time now a firm mathematical discipline, Statistical Decision Theory, has been applied to optimize signal reception in many important communication situations. Its most conspicuous success has been in the restricted problem of the reception of a (not too large) number of exactly known signals corrupted by additive independent gaussian noise with a known covariance function. Optimum reception in this case may be obtained through use of a correlation (or matched filter) receiver which uses replicas of the known signals stored either as waveforms or as the impulse responses of linear, time-invariant filters. When faced with the lack of specific knowledge of channel parameters, decision theory is less authoritative in specifying the form of the receiver. In this more difficult case, it is sometimes plausible to use an adaptive approach and estimate the possible waveforms which could exist at the receiver before noise is added, and to crosscorrelate the noisy received waveforms against these estimates. In fact, the application of decision theory to a channel which can be described by a multi-dimensional gaussian process with a zero mean and known covariance function, leads specifically to an adaptive receiver which generates reference waveforms from such estimates.



$$y(t) = \int_0^t x(\tau) h(t, \tau) d\tau + n(t)$$

*Generalized model of a communication channel.*


At the Philco Research Center, scientists are continuing to study adaptive techniques. Typical programs include 1.) the problems of communicating over channels whose statistics may be both poorly known and changing; and 2.) investigation of non-parametric methods of statistical inference which are not greatly sensitive to unknown parameters.

The Research Division offers attractive staff positions to scientists in a number of areas in addition to communication theory. Electrodynamics, plasma physics, ultramicrowaves, optical communications, and the physics of materials are among the areas of current interest. Letters of application or inquiry may be addressed to:

*Mr. W. M. Diefenbach  
Personnel Director  
Philco Research Center  
Blue Bell, Penna.*

RESEARCH CENTER

**PHILCO**

 Famous for Quality the World Over®

*An equal opportunity employer*

---

for placement ●  
or ■ personnel

data processing • mathematical  
sciences • operations research  
all levels • coast to coast  
contact us in confidence

**HERBERT HALBRECHT**

332 south michigan ave •  
chicago four, illinois •  
suite 540 harrison 7-2876

**MANAGEMENT COUNSEL  
EXECUTIVE RECRUITMENT**



#### **HANDBOOK OF STATISTICAL TABLES**

By DONALD B. OWEN, *Sandia Corporation*

Offers an unusually complete collection of tables of functions used in statistics. Extremely helpful to statisticians, quality control and industrial engineers. Some 100 different tables are included.

c. 576 pp, 1962—\$10.00

#### **MATHEMATICAL PROGRAMMING**

By S. VAJDA, *British Admiralty Research Laboratory, Teddington*

This well-organized book offers a large number of fine worked-out examples and problems. Early chapters offer an exhaustive treatment of the theoretical foundations, always with the use of fairly elementary mathematics. Later chapters guide the reader to the rapidly expanding frontiers of the subject.

310 pp, 77 illus, 1961—\$8.50

#### **INDEX OF MATHEMATICAL TABLES—Second Edition**

By A. FLETCHER AND J. C. P. MILLER, L. ROSENHEAD AND L. J. COMRIE, *Scientific Computing Service, Ltd.*

Indexes the contents of all important mathematical tables published since Briggs' *Logarithmorum Chilias Primae* (1617). Some previously unpublished tables included.

2 Vols. Dec. 1961—approx. \$42.00

..... New from Addison-Wesley!  
Examination copies gladly provided.  
Write: 519 South Street  
Reading, Massachusetts



## Guide to Tables in Mathematical Statistics

By J. R. GREENWOOD and H. O. HARTLEY

This *Guide*, the most comprehensive compilation available, is the product of more than twenty years' work under the sponsorship of the National Research Council, produced by and for the Committee on Mathematical Tables and Aids to Computation. The Committee believes that this monumental work will be a great service to the mathematical and statistical community, that scientists and technologists in many fields will find this *Guide* a valuable index to the widely scattered tables of probability and statistics needed in their work. 1076 pages. \$8.50

Order from your bookstore, or

PRINCETON UNIVERSITY PRESS  
Princeton, New Jersey

## INTERNATIONAL JOURNAL OF ABSTRACTS STATISTICAL THEORY AND METHOD

A Journal of the International Statistical Institute

The aim of this journal of abstracts is to give complete coverage of published papers in the field of statistical theory (including associated aspects of probability and other mathematical methods) and new published contributions to statistical method.

All contributions in the following five journals—being wholly devoted to this field—are abstracted: *Annals of Mathematical Statistics*; *Biometrika*; *Journal, Royal Statistical Society (Series B)*; *Bulletin of Mathematical Statistics*; *Annals, Institute of Statistical Mathematics*; and a further group of six journals are abstracted on a virtually complete basis as follows: *Biometrics*; *Metrika*; *Metron*; *Review, International Statistical Institute*; *Technometrics*; *Sankhyā*. There are about 250 other journals partly devoted to statistical theory and method from which the appropriate papers are abstracted.

The abstracts are about 400 words long—the recommendation of UNESCO for the “long” abstract service: they are in the English language although the original language of the paper is noted on the abstract together with the name of the abstractor. In addition the address of the author(s) are given in detail to facilitate contact in order to obtain further detail or request an off-print. The journal is published quarterly and contains approximately 1000 abstracts per year.

A scheme of classification has been developed for the abstracts that is flexible and facilitates the transfer of code numbers to punched cards. A unique aspect of this journal is that the pages are colour-tinted according to the main sections of classification. This method of colour-coding the pages provides a distinctive and powerful visual aid in the identification of abstracts in whatever manner the journal is filed for reference.

Annual Subscription	£5 (U.S.A. and Canada \$16.00)
Single Number	30s. (U.S.A. and Canada \$4.50)

OLIVER AND BOYD LTD.

Tweeddale Court, 14 High Street, Edinburgh, 1



# BIOMETRIKA

Volume 48, Parts 3 and 4

Contents

December 1961

## Memoirs:

DARNOCH, J. N. The two sample capture-recapture census when tagging and sampling are stratified. FRASER, D. A. S. The fiducial method and invariance. WETHERILL, G. B. Bayesian sequential analysis. BROWN, BYRON WM. (Jr.) Some properties of the Spearman estimator in bivariate. SLATER, PATRICK. Inconsistencies in a schedule of paired comparisons. MORGENTHAU, GEORGE W. Some circular coverage problems. BARTHOLOMEW, D. J. Ordered tests in the analysis of variance. YOUNG, D. H. Quota fulfilment using unrestricted random sampling. WALKER, A. M. Large-sample estimation of parameters for moving-average models. REIERST, OLAV. Linear and non-linear multiple comparisons in logit analysis. SAW, J. G. Estimation of the normal population parameters given a type I censored sample. QUESENBERY, C. P. AND DAVID, H. A. Some tests for outliers. KINGMAN, J. F. C. The ergodic behaviour of random walks. KAHIRAGAR, A. M. The goodness-of-fit of a single (non-isotropic) hypothetical principal component. JOHN, S. On the evaluation of the probability integral of the multivariate  $t$ -distribution. IMHOFF, J. P. Computing the distribution of quadratic forms in normal variables. CHANDA, K. C. Comparative efficiencies of methods of estimating parameters in linear autoregressive schemes. MOSTELLER, FREDERICK AND YOUTZ, CLEO. Tables of the Freeman-Tukey transformations for the binomial and Poisson distributions.

Miscellaneous: Contributions by J. D. BIGGERS; E. J. BURR AND GWENDA CANE; C. F. CROUSE; P. D. FINCH; GERALD J. GLASSER AND ROBERT F. WINTER; D. HOGDEN, R. S. PINKHAM AND M. B. WILK; B. K. KALE; J. G. SAW; A. ZINGER.

## Reviews

## Other Books Received

## Corrigenda

The subscription, payable in advance, is 54/- (Or \$8.00) per volume (including postage). Cheques should be made payable to *Biometrika*, crossed "a/c *Biometrika Trust*" and sent to The Secretary, *Biometrika Office*, University College, London, W.C.1. All foreign cheques must be drawn on a Bank having a London agency.

Issued by THE BIOMETRIKA OFFICE, University College, London.

## Privileges of Membership in the AMERICAN MATHEMATICAL SOCIETY

### FREE Subscriptions to:

NOTICES OF THE AMERICAN MATHEMATICAL SOCIETY, published 7 times a year, includes programs of the meetings of the Society, along with news items, announcements, personal items, and abstracts of papers presented at meetings.

BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY, publishes research announcements of timely interest, texts of invited addresses and book reviews; issued bimonthly.

PROCEEDINGS OF THE AMERICAN MATHEMATICAL SOCIETY, publishes original research results; issued bimonthly.

Available instead of the PROCEEDINGS upon payment of \$10.00  
in addition to dues:

TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY  
OR MATHEMATICAL REVIEWS

### REDUCED rates for members to:

TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY

(List price, \$32.00) \$16.00 per year

MATHEMATICAL REVIEWS

(List price, \$50.00) \$16.00 per year

ALL BOOKS PUBLISHED BY THE SOCIETY available at 25% discount

NINE JOURNALS published by other Institutions 20%-45% discount

For full information write to:

**THE AMERICAN MATHEMATICAL SOCIETY**

190 Hope Street, Providence, Rhode Island

Dues: \$14.00 per year

# JOURNAL OF THE ROYAL STATISTICAL SOCIETY

*Series B (Methodological)*

Single Parts 40s

Ann. Sub. incl. postage  
£3 2s 0d U.S. \$9

## CONTENTS OF VOL. 23, PART 2, 1961

- A Test of Homogeneity of Means under Restricted Alternatives (With Discussion). D. J. BARTHOLOMEW  
Optimal Programmes for Varietal Selection (With Discussion). R. N. CURNOW  
Some Stochastic Processes with Absorbing Barriers. T. L. SAATY  
Some Sequential Tests using Range. W. G. GILCHRIST  
First Emptiness in a Finite Dam. B. WEESAKUL  
The Construction of Optimal Designs for the One-Way Classification analysis of variance. J. AITCHISON  
Some Simple Wear-Dependent Renewal Processes. A. MENGER  
An Exact Distribution of the Behrens-Fisher-Welch Statistic for Testing the Difference between the Means  
of Two Normal Populations with Unknown Variances. W. D. RAY & A. E. N. T. PITMAN  
Closed Queueing Systems: A Generalisation of the Machine Interference Model. F. BENSON & G. GREGORY  
Testing for a Jump in the Spectral Function. E. J. HANNAN  
Moments of a Truncated Bivariate Normal Distribution. S. ROSENBAUM  
On Quadratic Estimates of the Interclass Variance for Unbalanced Designs. R. S. READ  
Mixtures of Geometric Distributions. H. E. DANIELS  
Prediction by Exponentially Weighted Moving Averages and Related Methods. D. R. COX  
The Randomisation Bases of the Problem of the Annalagimation of Weighted Means. D. E. BARTON & C. L.  
MALLOWS  
Estimation from Quantiles in Destructive Testing. H. P. STOUT & F. STERN  
A Note on Maximum-Likelihood in the Case of Dependent Random Variables. S. D. SILVEY  
Some Properties of the Asymptotic Variance of the Sample Quantiles and Mid-Ranges. P. K. SEN  
Similarities between Likelihoods and Associated Distributions A Posteriori. D. A. SPROTT  
A Method of Maximum Likelihood Estimation. F. S. G. RICHARDS  
The Intervals between Regular Events Displaced in Time by Independent Random Deviations of Large  
Dispersion. T. LEWIS  
An Approach to the Scheduling of Jobs on Machines. E. PAGE

The Royal Statistical Society, 21, Bentinck Street, London, W. 1

# SANKHYĀ

*The Indian Journal of Statistics*

Contents of Series A, Vol. 23—Part 3

- An elementary theory of the Brownian motion random function..... J. Kampé De Fériet  
On testing the mean of a discrete linear process..... K. R. Parthasarathy  
On a problem of Bartholomew in life testing..... P. S. Swamy and S. A. D. C. Doss  
Studentization of two-stage sample means from normal populations with unknown common variance  
Harold Ruben  
Testing for the single outlier in a regression model..... K. S. Srikantan  
Minimal sufficient statistics for the balanced incomplete block design under an Eisenhart model II  
David L. Weeks and Franklin A. Graybill  
Asymptotic bias and variance of ratio estimates in generalized power series distributions and certain applications..... G. P. Patil  
Method of constructing two mutually orthogonal latin squares of order  $3n + 1$ ..... P. Kesava Menon  
Combinatorial arrangements analogous to orthogonal arrays..... C. Radhakrishna Rao  
On the structure and combinatorial properties of certain semi-regular group divisible designs..... W. S. Saraf  
A factorial approach to construction of true cost of living index and its application in studies of changes in  
national income..... K. S. Banerjee  
Generation of random permutations of given number of elements using random sampling numbers  
C. Radhakrishna Rao

The subscription price per volume for each of the series is: USA and Canada: \$10.00; Other Countries: £3.  
The subscription price per volume for both the series taken together is: USA and Canada: \$18.00; Other Countries: £5-6s.

Subscription is payable in advance to:

The Secretary, Statistical Publishing Society,  
204/1 Barrackpore Trunk Road,  
Calcutta-35, India.

# ESTADISTICA

Journal of the Inter American Statistical Institute

Vol. XIX, No. 70

March 1961

## CONTENTS

- Las Conferencias Interamericanas de Estadística.....César A. Orantes  
La Labor del IASI en el Campo de Organización y Administración Estadística  
Alfonso Perea Posada  
La Formación del Personal Estadístico.....Tulo H. Montenegro  
About the Standardization of Some Types of Demographic Studies  
João Lyra Madeira  
El Uso de las Estadísticas en la Formulación y Evaluación de los Programas Sociales  
Octavio Cabello  
Aumento en el Alquiler de las Viviendas en Estados Unidos entre 1940 y 1950 (traducción).....Margaret G. Reid  
Special Feature: Principales Programas Estadísticos Federales de los Estados Unidos.  
Legal Provisions. Institute Affairs. Statistical News. Publications.

*Published quarterly*      *Annual subscription price \$3.00 (U. S.)*

INTER AMERICAN STATISTICAL INSTITUTE

Pan American Union

Washington 6, D. C.

# TRABAJOS DE ESTADISTICA

Review published by the "Instituto de Investigaciones Estadísticas" of the "Consejo Superior de Investigaciones Científicas" Madrid-6, Spain.

VOL. XII.

Cuad. I and II.

## CONTENTS

- H. BRENY.—Un théorème de la théorie statistique des modèles linéaires.  
M. DWARS.—The distribution of linear combinations of random divisions of an interval.  
M. M. RAO.—Some remarks on independence of statistics.  
S. RÍOS, J. BEJAR and J. M. GARCIA.—Un problema de normalización de medidas como base para un problema de producción e inventario.  
J. C. SMIT.—Estimation of the mean of a stationary stochastic process by equidistant observation  
S. RÍOS.—Procesos de decisión.

### NOTAS.

- S. RÍOS.—La investigación operativa y los métodos clásicos de organización.  
J. M. GARCIA.—La investigación operativa en la industria química.  
I. YARZEL.—Nota sobre programación lineal estocástica.

CRONICA • BIBLIOGRAFIA • CUESTIONES PROPUESTAS • EJERCICIOS.

For everything in connection with works, exchanges and subscription write to Professor Sixto Ríos, Instituto de Investigaciones Estadísticas, C.S.I.C., Serrano 123, Madrid-6, Spain.

The Review is composed of three fascicles published three times a year (about 350 pp), and its annual price is 100 pesetas for Spain and 240 pesetas for all other countries.

# JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Volume 56

December, 1961

Number 296

Rectifying Inspection of Lots.....	F. J. Anscombe
On Comparing Intensities of Association between Two Binary Characteristics in Two Different Populations.....	Agnes Berger
Occupational Components of Educational Differences in Income.....	Otis Dudley Duncan
Failure of Enumerators to Make Entries of Zero: Errors in Recording Childless Cases in Population Censuses.....	M. A. El-Badry
Note on the Missing Plot Procedure in a Randomized Block Design.....	Leroy Folks and Del Lon West
A Bivariate Extension of the Exponential Distribution.....	John E. Freund
Stepwise Least Squares: Residual Analysis and Specification Error.....	Arthur S. Goldberger
Statistical Methods for the Mover-Stayer Model.....	Leo A. Goodman
Gamma Distribution in Acceptance Sampling Based on Life Tests.....	Shanti S. Gupta and Phyllis A. Groll
On the Resolution of Statistical Hypotheses.....	Robert V. Hogg
A Simple Theoretical Approach to Cumulative Sum Control Charts.....	Norman L. Johnson
A Nomograph for Computing Partial Correlation Coefficients.....	Keith W. Lee and Frederic M. Lord
Non-Additivity in Two-Way Analysis of Variance.....	John Mandel
The Statistical Analysis of Industry Structure: An Application to Food Industries.....	Lee E. Preston and Earl J. Bell
The Asymptotic Variances of Method of Moments Estimates of the Parameters of the Truncated Binomial and Negative Binomial Distributions.....	S. M. Shah
Forecasting Industrial Production.....	H. O. Stekler
Residence Histories and Exposure Residences for the United States Population.....	Karl E. Taeuber, William M. Hoensel and Monroe G. Sirken
Testing the Independence of Regression Disturbances.....	Henri Theil and A. L. Nagar

For further information, please contact:

## AMERICAN STATISTICAL ASSOCIATION

1757 K Street, N.W.

Washington 6, D. C.

# ECONOMETRICA

*Journal of the Econometric Society*

*Contents of Vol. 29, No. 3 - July 1961*

L. R. KLEIN: A Model of Japanese Economic Growth, 1878-1937
MICHAEL C. LOVELL: Manufacturers' Inventories, Sales Expectations, and the Acceleration Principle
JOHN F. MUTH: Rational Expectations and the Theory of Price Movements
LESLIE KISH: Efficient Allocation of a Multi-Purpose Sample
FREDERICK V. WAUGH: The Place of Least Squares in Econometrics
YEHUDA GRUNFELD: The Interpretation of Cross Section Estimates in a Dynamic Model
RAGNAR FRISCH: A Reconsideration of Domar's Theory of Economic Growth
DENIS SARGAN: The Maximum Likelihood Estimation of Economic Relationships with Auto-Regressive Residuals
PETER E. DE JANOSI: The Statistical Discrepancy in the National Accounts Revisited
E. MALLINVAUD: The Estimation of Distributed Lags: A Comment
REPORT OF THE STANFORD MEETING
REPORT OF THE ST. LOUIS MEETING
BOOK REVIEWS
ANNOUNCEMENTS AND NOTES



THE INSTITUTE OF MATHEMATICS

Abstract of

Contributed or Invited Paper

Name with title, institution, and address of author:

Title of Paper:

Meeting to which paper pertains (Contributed papers by title need not pertain)

Check status of paper: Contributed ☐ Invited ☐

Check method of presentation: In person ☐ By title ☐

If slide or projection equipment is wanted, describe it clearly. A specific kind of

Journal to which paper will be offered:



**STATISTICAL STATISTICS**

**ed Paper**

ertain to any meeting; in that case leave this space blank);

kind of equipment cannot be assured.

Journal to which paper will be offered:

TWO COPIES of an abstract of every contributed paper for presentation at the time designated in the announcement of the meeting. Meetings are announced in *Mathematical Statistics*. The title of a paper not in form for publication should be

Abstracts for invited papers are optional.

Anyone who is not a member of the Institute must be introduced by a member by title. A signed statement from the introducing member should be attached to the abstract.

Papers may not be submitted if published in full before the date of the meeting by the society.

Only one paper per member may be presented in person at any one meeting. The number of papers presented by title is not restricted.

Abstracts of invited papers at joint sessions with other organizations should be arranged for their publication. Contributed Paper sessions are usually announced in a special announcement will be made about publication of abstracts.

Abstract blanks will be furnished to members by the Secretary on request in order to save time and reduce error. Abstracts may be returned to their authors without giving full information.

on at a meeting of the Institute must be sent to the Editor by  
announced by mail and on the back cover of the Annals of  
n should end with the words "Preliminary Report."

by a member in order to present a paper, either in person or  
ached to the abstract.

the Institute meeting or if previously presented to any learned

ne meeting. Ten minutes is the normal time allotted to each.

ns should not be submitted to the Editor if the other organiza-  
usually not held jointly with other organizations. When they are,

request. Members are urged to use the printed abstract blanks in  
eir authors and publication delayed if they are not clear or do

## INSTRUCTIONS TO AUTHOR

The abstract should state clearly the methods and results of the paper within 200 words, or the equivalent, and in general should consist of a single paragraph of the abstract and should follow the standard format of the *Annals of Mathematics*. Letters and formulas should be expressed as simply as possible; for example, formulae should be written in horizontal sequence. Displayed formulas should be avoided whenever possible. Abstracts should be double spaced, and in form for immediate publication. Remember to send two copies.

Indicate marginally in pencil the names of Greek, German, or script letters, and symbols to a minimum. Also indicate names of symbols that may be confused with capital oh, small oh (when used in formulas) and zero. Distinguish between capital

Author: these columns  
for instructions to  
printer.

Author: type abstract below

## AUTHORS

with as little detail as possible. Abstracts should not exceed one paragraph. References should be contained in the body of the text. Unusual symbols should be avoided, and formula symbols should, whenever feasible, run in a horizontal line. Abstracts should be typewritten to the maximum extent, and two copies.

Letters and those of unusual symbols. Keep the use of such symbols may be ambiguously read, for example one and el, or in capital and lower case forms when they look alike.

---

below





# THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

## OFFICERS

### *President:*

A. H. Bowker, Department of Statistics, Stanford University, Stanford, California

### *President-Elect:*

T. W. Anderson, Department of Mathematical Statistics, Columbia University, New York 27, New York

### *Secretary:*

G. E. Nicholson, Jr., Department of Statistics, University of North Carolina, Chapel Hill, North Carolina

### *Treasurer:*

Gerald Lieberman, Institute of Mathematical Statistics, Sequoia Hall, Stanford University, Stanford, California

### *Editor:*

J. L. Hodges, Jr., Department of Statistics, University of California, Berkeley 4, California

The purpose of the Institute of Mathematical Statistics is to encourage the development, dissemination, and application of mathematical statistics.

Membership dues including a subscription to the ANNALS OF MATHEMATICAL STATISTICS are \$10.00 per year for residents of the United States or Canada and \$5.00 per year for residents of other countries. There are special rates for students and for some other classes of members. Inquiries regarding membership in the Institute should be sent to the Treasurer of the Institute.

Membership in the Institute of Mathematical Statistics is not required of authors of papers in the *Annals of Mathematical Statistics*, and the fact of membership or non-membership is given no weight in the consideration of submitted manuscripts.

## Contents (Continued)

On a Special Class of Recurrent Events.....	M. P. SCHÜTENBERGER	1201
Maximum Likelihood Characterization of Distributions.....	HENRY TEICHER	1214
On the Distribution of First Significant Digits.....	ROGER S. PINKHAM	1223
Markov Renewal Processes: Definitions and Preliminary Properties		
	RONALD PYKE	1231
Markov Renewal Processes with Finitely Many States.....	RONALD PYKE	1243
A Convexity Property in the Theory of Random Variables Defined on a Finite Markov Chain.....	H. D. MILLER	1260
Limit Distribution in the Theory of Counters.....	G. SANKARANARAYANAN	1271
The Transient Behavior of a Single Server Queuing Process with Recurrent Input and Gamma Service Time.....	LAJOS TAKÁCS	1286
Efficient Estimation of a Regression Parameter for Certain Second Order Processes.....	CHARLOTTE T. STRIEBEL	1299
Two Similar Queues in Parallel.....	J. F. C. KINGMAN	1314
Queues with Batch Departures I.....	F. G. FOSTER AND K. M. NYUNT	1324
<b>NOTES</b>		
On the Chapman-Kolmogorov Equation.....	JACK KARUSH	1333
A Generalization of a Theorem of Balakrishnan		
	N. DONALD YLVISAKER	1337
The Opinion Pool.....	M. STONE	1339
Correction Notes and Acknowledgments of Priority.....		1343
Abstracts of Papers.....		1346
News and Notices.....		1353
Report of the President for 1961.....		1358
Report of the Editor for 1961.....		1363
Publications Received.....		1365

## MEETINGS OF THE INSTITUTE

### TENTATIVE SCHEDULE

**EASTERN REGIONAL MEETING—New York City, December 27-30, 1961**

**EASTERN REGIONAL MEETING—  
Chapel Hill, North Carolina, April 12-14, 1962**

**WESTERN REGIONAL MEETING—Albuquerque, New Mexico,  
April 19-20, 1962**

**EUROPEAN REGIONAL MEETING—  
Dublin, Ireland, September 3-5, 1962**

**ANNUAL MEETING—Minneapolis, Minnesota, September 7-10, 1962**

Abstracts should be submitted in duplicate to the Editor, on abstract blanks, which can be found in the back of every *Annals*, beginning with the March 1961 issue, or can be obtained from the IMS Secretary. Abstracts must be received at least 50 days before the first day of the meeting at which they are to be presented, indicating whether presented by title or in person. (Only one contributed paper may be given in person at any one meeting.) They may be printed prior to the publication of the report of the meeting. Those received by April 30 will appear in the September *Annals*, by July 31 in December, etc. Abstracts should be limited to 200 words or the equivalent, and should avoid displayed expressions and complicated formulae. They can be accepted from non-members of the IMS only if introduced by members. Abstracts must follow the stylistic requests on the abstract blank, or they may be returned.

